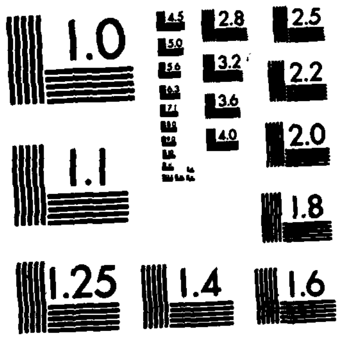
AD-A124 225   SPEECH ENVELOPE NORMALIZATION A METHOD TO IMPROVE SNR   1/2
AND SUPPRESS NOISE I..(U) NATIONAL RESEARCH COUNCIL
WASHINGTON DC COMMITTEE ON VETERINA.. J C SPRINGETT
UNCLASSIFIED   DEC 82 CECOM-82-J066-F DAAB07-82-C-J066   F/G 17/2   NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

RESEARCH AND DEVELOPMENT TECHNICAL REPORT
CECOM

REPORT NO. CECOM-82-J066-F

SPEECH ENVELOPE NORMALIZATION, A METHOD TO IMPROVE SNR AND SUPPRESS NOISE IN PRESENT AND FUTURE RADIO SYSTEMS

JAMES C. SPRINGETT

NEOCOMM SYSTEMS, INC.
4529 ANGELES CREST HWY., SUITE 204A
LA CANADA-FLINTRIDGE, CA 91011

DECEMBER 1982

FINAL REPORT FOR PERIOD 15 MARCH TO 15 SEPT 1982

AUTHORIZED FOR PUBLIC RELEASE; UNLIMITED DISTRIBUTION

CECOM ATTENTION: DRSEL-COM-RN-3
U S ARMY COMMUNICATIONS-ELECTRONICS COMMAND
FORT MONMOUTH, NEW JERSEY 07703

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS
BEFORE COMPLETING FORM

1. REPORT NUMBER
CECOM-82-J066-F

2. GOVT ACCESSION NO.
AD-A124225

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Speech Envelope Normalization,
A Method To Improve SNR and Suppress Noise
in Present and Future Radio Systems

5. TYPE OF REPORT & PERIOD COVERED
Final Report
15 March 82 to 15 Sept. 82

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
James C. Springett

8. CONTRACT OR GRANT NUMBER(s)
DAAB07-82-C-J066

9. PERFORMING ORGANIZATION NAME AND ADDRESS
NeoComm Systems, Inc.
4529 Angeles Crest Highway, Suite 204A
La Canada-Flintridge, CA 91011

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
1L1 612701 AH92

11. CONTROLLING OFFICE NAME AND ADDRESS
U. S. Army CECOM
ATTENTION: DRSEL-COM-RN-3
Fort Monmouth, NJ 07703

12. REPORT DATE
December, 1982

13. NUMBER OF PAGES
137

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Authorized for public release; unlimited distribution

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Speech Processing, Speech Compression, Speech Envelope Normalization, SNR Improvement

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Speech envelope normalization (EN) is a compression process that transforms dynamic speech waveforms into constant envelope signals which are optimum for transmission and reception using analog radio equipment. The EN process reduces dynamic range requirements, maximizes speech SNR, increases intelligibility, and suppresses noise. The Report covers EN theory and properties, EN circuit implementations, experimental evaluations, and applications analysis.

NOTICES

Disclaimers

The citation of trade names and names of manufacturers in this report is not to be construed as official Government indorsement or approval of commercial products or services referenced herein.

Disposition

Destroy this report when it is no longer needed. Do not return it to the originator.

ACKNOWLEDGEMENTS

The success of the efforts described in this report have been largely dependent upon the contributions of several individuals. Dr. Marvin K. Simon provided most of the supporting analysis and aided the author and principal investigator in developing some of the key concepts. Thanks also are given to Dr. Paul Richter for stimulating discussions on speech properties. Mr. Joe T. Sumida was responsible for the bulk of circuit designs used for the proof-of-concept and subsequent Demonstrator units. Mr. Jeffery C. Packard functioned as technician and draftsman, and provided the physical design for the Demonstrator. Finally, Ms. Marion T. Ingham is to be commended for editing and typing of this final report.

SUMMARY

Speech envelope normalization (EN) is a form of signal amplitude compression which results in a speech waveform that has a constant (i.e., flat) envelope. In other words, the speech signal following EN appears much the same as if the speech had been passed through a hard limiter (or infinite peak clipper). However, unlike the hard limiter, the EN speech waveform does not contain large in-band harmonics, and the EN process is completely reversable. This latter property means that speech may be converted to its EN form for transmission over a communication link, and at the receiving end the EN waveform can be converted back or expanded to obtain the original speech signal.

The EN speech signal is optimum for transmission and reception using analog radio equipment (e.g., AM, FM, SSB, etc.) because it (1) minimizes the dynamic range requirements on the bulk of the transmitter and receiver electrical circuits, (2) maximizes the radio transmitter modulation index, thereby maximizing the received speech signal-to-noise ratio (SNR), (3) increases speech intelligibility, and (4) suppresses receiver electrical noise as a result of the expansion operation. Additionally, the EN speech form is capable of raising articulation levels when listening must be performed within high acoustic noise environments. As a result, EN speech processing can provide more reliable communications in high-noise (electrical and acoustical) situations, reduce listening fatique, and generally enhance the ability to receive spoken messages without misunderstanding. Additional advantages which may accrue from the use of EN, especially in the design of new radio equipment, are more RF channels for a fixed total band allocation, and reduction of the RF transmitter power requirement.

The production of a high quality EN speech signal requires a certain amount of specialized processing, a principal requirement being that of deriving from the original speech signal its true envelope. EN is then obtained by dividing the original speech waveform by this

envelope. In order to effect expansion at the receiver, the speech envelope must be transmitted over the communication link along with the EN speech. Additionally, it is necessary that these operations be compatible with existing radios and their limitations (especially their passband characteristics). For these reasons a research program was undertaken to (1) study the theoretical properties of EN signals, (2) design and evaluate circuits for EN signal production, (3) perform measurements to verify EN characteristics and system performance, and (4) construct a EN demonstrator that can be used for proof of concepts and subjective evaluations. Results of these activities are detailed in this Report, and highlights of the results, conclusions, and recommendations are summarized forthwith.

A firm theoretical basis for the EN process and its properties was derived based upon a technique long used for single sideband speech transmission known as "RF clipping." Analysis, in conjunction with the properties of the Hilbert Transform and the theory of analytic signals, established the mathematical form of the true speech envelope, and the fact that a constant envelope signal results when the original speech signal is divided by its true envelope. Additional theoretical studies determined temporal and spectral characteristics of EN signals.

Expected application improvements from the use of EN processing were studied. Direct SNR increases of between 8 dB and 15 dB will typically be obtained (the exact value being conditional upon existing radio link parameters). Subjective SNR improvements resulting from receiver expandor operation should range between 15 dB to 30 dB (depending upon circumstances). As an alternative to direct SNR increases, savings in channel bandwidth or transmitter power can be obtained by using EN. For a typical FM link, the channel bandwidth can be decreased from 30 KHz to 10 KHz, or the transmitter power reduced by a factor of six.

Since calculation of the Hilbert transform of the speech signal is essential to the formation of the true speech envelope, several basic methods, and a number of circuit implementations were studied, breadboarded, and evaluated. Two different approaches emerged, one employing cascaded active all-pass networks, and the other using a charged coupled device sampled data delay line to realize the transversal filter form of the Hilbert transform integral. Both were incorporated into the EN demonstrator. Digital configurations were considered, but judged too complex to be competitive with the other approaches. Several circuit designs were also examined for implementing the operations required to form the envelope signal, and perform speech by envelope signal division. The final design has an operating dynamic range of 54 dB.

A property of the temporal pattern of speech is that there are many short segments when the waveform is zero (between words, during pauses, etc.). These segments are referred to as silence periods. Since noise of some form always accompanies the speech, the EN circuits (because of their large dynamic range) will greatly amplify this noise during speech silence. Under certain circumstances this byproduct of EN could prove detrimental. As a result, a speech vs. silence detector was designed so that the output of the EN processor may be disabled whenever only noise is present. This operation works well for low noise levels, but becomes somewhat marginal for large noise because of the need to accurately set a decision threshold between the speech and noise levels.

Transmission of the speech envelope is accomplished by modulating it onto a subcarrier or pilot above the upper frequency limits of the EN speech signal. After studying the relative advantages and disadvantages of several types of pilot modulation, FM was selected. Performance analysis established that the pilot component amplitude needs to be between 10 dB and 6 dB below the EN speech signal amplitude for acceptable operation.

A large amount of experimental data was obtained on EN system performance. It was determined that the speech envelope contains frequency components well up into the hundreds of Hz range which are essential to quality

EN waveform production. However, for expansion, the envelope may be lowpass filtered to as low as 100 Hz without seriously degrading performance. Modification of the speech frequency spectrum by the EN process is small, especially in terms of bandwidth expansion. Therefore, the EN waveform may be lowpass filtered to the original speech signal's band limit without significantly affecting its constant envelope nature. Voiced or vowel sounds are those most dramatically affected by the EN operation. If it is desired to listen directly to the EN speech (i.e., eliminate expansion), it was found that preemphasis of the original speech spectrum by a rising 6 dB/octive characteristic prior to EN greatly increases the intelligibility of the voiced sounds. Other measurements revealed heretofore unknown properties of the true speech envelope.

Overall, the research program achieved every major planned goal, and the proof of concept is complete. The only real disappointments were that a valid method was not devised for measuring expandor subjective SNR improvement, and time did not permit testing of the EN demonstrator with an actual radio link. However, future effort appears fully justified, and should be directed toward (1) improving speech vs. silence detector performance, (2) additional study of envelope linking, (3) quantitative testing of EN speech system articulation using a jury of listeners, (4) design and construction of prototype EN circuits for use with existing radio sets, and (5) field testing of EN radio equipment.

CONTENTS

PAGE

1.0 INTRODUCTION

This Report describes a research effort to develop and evaluate a technique known as speech envelope normalization (EN), which is capable of increasing voice signal-to-noise ratio (SNR) and improving speech intelligibility when the speech is transmitted and received using analog radio links. The overall degree of enhancement depends upon a number of complex factors, some of which are objective (capable of being directly measured), while the others are subjective (requiring perceptive opinion). Additionally, the use of EN increases the efficiency of radio links in terms of the usual measures of transmitter RF power and signal bandwidth.

The overall objective of the reported research is to investigate speech envelope normalization from several perspectives, including theory, analytical and experimental performance, circuit mechanizations, and radio link designs. A brief background on speech intelligibility as affected by radio links is given in Section 2.0. Section 3.0 introduces the principals and properties of EN, including a historical motivation for the effort. Section 4.0 analyzes the speech transmission and radio link performance improvements that may be expected with EN.

In Sections 5.0 and 6.0, various direct methods and circuits which function to derive the true speech envelope and perform envelope normalization are discussed. Performance and complexity tradeoffs are made, and the practical problem of discriminating against noise when speech is absent is investigated. Section 7.0 then outlines other possible EN signal production approaches, and explains why they are either ineffective, or unduly complex, relative to the direct methods.

Since EN signal expansion requires knowledge of the envelope waveform at the receiver, methods for linking the envelope are examined and analyzed in Section 8.0. Section 9.0 then presents the experimental results obtained with the Breadboard Model EN compandor, including the temporal performance, speech spectrum alteration, and subjective effects under a variety of configuration and noise conditions.

In Section 10.0 the Breadboard is described in terms of its construction, features, and capabilities. Finally, Section 11.0 presents conclusions based on the research results, and makes recommendations for future work.

2.0 BRIEF BACKGROUND ON SPEECH INTELLIGIBILITY AS AFFECTED BY RADIO LINKS

Any system that involves the transmission of voice signals must contend with a wide range of differing inputs, i.e., the high variable nature of the speaking population. The voice dynamic range of a given speaker or talker when engaged in telephonic or radio communication is typically 20 dB, while the dynamic range over the total population of speakers, from the softest to the loudest is some 30 dB. Thus, the voice link must be able to efficiently accommodate an overall input dynamic range of 50 dB. A second very important consideration is the quality of the speech reproduced at the output of the radio system. Significant measures in this regard are articulation or intelligibility, signal-to-noise ratio (SNR), distortion, and speaker identification. From the communication engineer's perspective, nothing could be less desirable than the aforementioned characteristics. Beset by pragmatic matters, such as available transmitter power constraints, channel bandwidth limitations, and system dictums such as "maximize the number of channels," "be compatible with interfacing and tandem systems," and "let's use the most advanced techniques available," the task of specifying any radio modulation system to efficiently handle speech becomes arduous.

The speech voltage waveform at the output of a microphone is very dynamic. Spoken words and phrases may be viewed as short bursts, between which the microphone output is essentailly zero. Figure 1 is such a characterization. For the immediate discussion each burst may be represented in terms of its envelope, which may be thought of as a slowly varying function which undulates in accord with the speaks and valleys of the waveform as pictured in Figure 1.

The bursts shown as Figure 1 may be thought of as words, each word composed of syllables. Syllables generally fall into two classes, voiced and unvoiced. Voiced syllables are those associated with vocal cord vibration. The vowel letters are an example of voiced sounds. Unvoiced syllables or

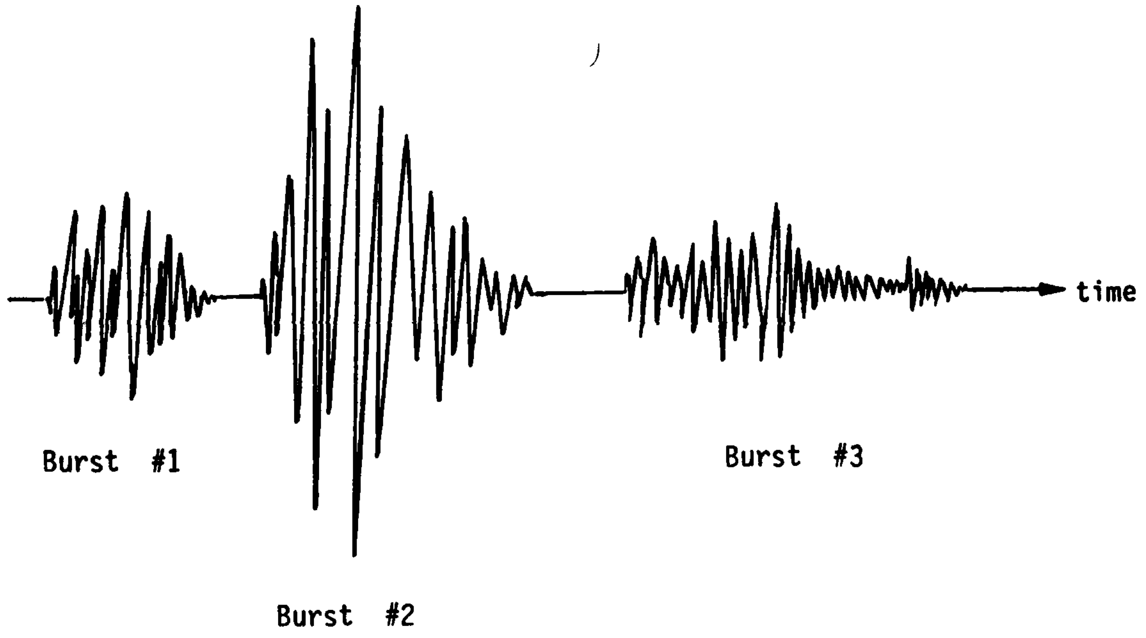


FIGURE 1 - TYPICAL SPEECH WAVEFORM

sounds are formed by forcing air though a constricted area of the vocal tract (especially in the mouth and lips) to produce air turbulence. Thus, the unvoiced sounds tend to be "noiselike" as compared to the voiced sound's periodic vibrations. Voiced syllables typically have a higher energy content than unvoiced syllables. Alternatively, the waveform envelope for voiced segments is considerably larger than that for unvoiced segments.

Now a pecularity of speech is that the unvoiced sounds are more critical to intelligibility than the voiced (especially vowel) sounds, yet they are weaker. Since all radio links add various forms of noise to the received speech signal, this noise tends to mask the weaker segments of the speech. As a result, the unvoiced sounds or weak syllables will have a relatively low SNR, which in turn will significantly reduce intelligibility or articulation. On the other hand, the SNR for the voiced sounds or strong syllables will be relatively high. Since the ear tends to filter-out all noise components except for small bands centered on the formant frequencies, the perceived average speech SNR may appear reasonably large, even though articulation has been compromised. Simply stated, unvoiced sounds have insufficient SNR while voiced sounds have an excess SNR.

Most analog speech radio systems are designed to accommodate the wide dynamic range of the speaking population, i.e., they are reasonably linear over a 50 dB range. For soft talkers who produce small electrical speech signals into the radio link transmitter, the linear system property results in correspondingly small electrical speech signals at the output of the radio link receiver. Because the additive noise level does not depend upon whether a talker is loud or soft, clearly the softer the speaker, the lower the received SNR, and the poorest intelligibility. If the overall link parameters are specified to meet a particular articulation index for the softest talkers, then the link will overperform for all other conditions. Such an approach is always wasteful of RF power and bandwidth.

These problems with speech transmission have been the concern of comunication engineers from the earliest days of telephony and radio. An approach for minimizing the differences between talkers is to employ a form of automatic gain or level control (AGC or ALC). Such circuits attempt to measure the average speaking level over several words, and adjust the amplification of a variable gain device so that the speech voltage is regulated at the output to a specific value. Although effective in reducing the talking population dynamic range from 30 dB to less than 10 dB, an ALC has little effect in altering the strong to weak syllable ratio.

A second operation applied to speech signals is that of compression, and acts to change the strong to weak syllable ratio. Although it is functionally similar to the ALC, a compressor estimates and changes the speech level at the syllabic rate. Thus, the averaging times involved with compression are on the order of one-tenth those used for ALC. As a result, the compressor attenuates the large segments and amplifies the weaker segments of the speech waveform. In effect, this is a form of speech signal distortion, which needs to be removed at the receiver if the speech is to sound perfectly natural. The reverse process is known as expansion, and together the compressor and expandor pair is known as a compandor. Compandors were used as early as the 1920's by the Bell Telephone System.[1] In some applications an ALC is used in tandem with a compressor in an attempt to combine the attributes of both devices, especially if the compression ratio (to be defined subsequently) is small.

[1] Fagen, M. D., Editor, A History of Engineering and Science in the Bell System, The Early Years (1875-1925). Bell Telephone Laboratories, Inc., 1975.

Companders operate according to a given compression/expansion ratio or law. Most laws are ν:1 in the logarithmic (dB) realm. The common telephone compressor has a 2:1 logarithmic input/output characteristic, that is, the output of the compressor changes 1 dB for every 2 dB of input signal change. Actually, it is an approximation to the signal's envelope that varies in this manner. Let the input speech signal v(t) be written as

$$v(t) = e(t)g(t), \tag{1}$$

where $e(t) \geq o$ is the envelope function, and g(t) represents the essential frequency components of the speech signal. The output signal from a 2:1 compressor is ideally given by

$$v_o(t) = \sqrt{e(t)}g(t). \tag{2}$$

ᵀn reality, this is never strictly realized because the usual mechanization, shown as Figure 2, is to employ an inexact estimate of e(t)

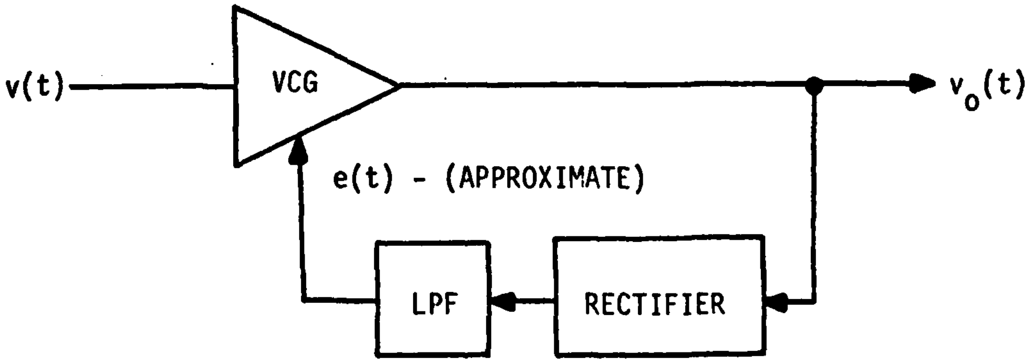


FIGURE 2 - 2:1 COMPRESSOR

(obtained using a rectifier followed by a lowpass filter) in conunction with a voltage or current controlled gain element in order to effect compression. Not only is the envelope estimate inexact, no delay of the input signal is made to compensate for the evelope estimator time constant, i.e., the delay introduced by the lowpass filter. A 4:1 compressor law has also found application, and is usually implemented by cascading two 2:1 compressors, with even less idealism.

In general, ideal ν:1 compression may be realized according to the relationship

$$v_0(t) = e(t)^{1/\nu} g(t). \qquad (3)$$

If the dynamic range of $v(t)$ is ηdB, then the dynamic range of $v_0(t)$ will be η/ν dB. Clearly, when $\nu=\infty$, the dynamic range of $v_0(t)$ must be 0 dB, and since

$$v_0(t) = g(t) \qquad (4)$$

under this condition, then $g(t)$ will also have an 0 dB dynamic range. This means that $g(t)$ must be a waveform which has a constant envelope. As a result, any speech signal may be viewed as a rapidly varying constant envelope waveform multiplied by a relatively slowly varying envelope function. Thus, if a means can be found to decompose speech into these components, then a more efficient, in fact optimum, radio link may be realized because the speech signal may, in effect, be infinitely compressed prior to transmission so that the strong to weak syllable ratio becomes unity. The process by which such is achieved is known as speech signal envelope normalization (EN).

Table 1 summarizes the performance of 2:1, 4:1, and EN compressors relative to the normal speaking population. The unaffected level is that which remains unchanged between compressor input to output, and is defined as the RMS level for the average speaker.

SPEAKING POPULATION	NO COMPRESSION	2:1 COMPRESSION	4:1 COMPRESSION	ENVELOPE NORMALIZATION
LOUD SPEAKER LEVEL				
Strong Syllable	25.0 dB	12.5 dB	6.25 dB	0.0 dB
RMS	15.0 dB	7.5 dB	3.75 dB	0.0 dB
Weak Syllable	5.0 dB	2.5 dB	1.25 dB	0.0 dB
AVERAGE SPEAKER LEVEL				
Strong Syllable	10.0 dB	5.0 dB	2.50 dB	0.0 dB
RMS (Unaffected Level)	0.0 dB	0.0 dB	0.00 dB	0.0 dB
Weak Syllable	-10.0 dB	-5.0 dB	-2.50 dB	0.0 dB
SOFT SPEAKER LEVEL				
Strong Syllable	-5.0 dB	-2.5 dB	-1.25 dB	0.0 dB
RMS	-15.0 dB	-7.5 dB	-3.75 dB	0.0 dB
Weak Syllable	-25.0 dB	-12.5 dB	-6.25 dB	0.0 dB

TABLE 1 - SPEAKING POPULATION SPEECH LEVELS
FOLLOWING IDEAL COMPRESSION

3.0 PRINCIPLES AND PROPERTIES OF SPEECH ENVELOPE NORMALIZATION

Throughout the Report, use will be made of the Hilbert Transform and its properties, the concept of analytic signals, and the mathematical definition of signal envelope. It is beyond the scope of this Report to propound these principals, and the unfamiliar reader may wish to resort to Reference 2 for Hilbert Transforms, References 2, 3, and 4 for analytic signals, and References 2 and 4 for the signal envelope.

3.1 Historical Perspective and Motivation

In his 1964 book[5] on single sideband (SSB), Pappenfus in discussing speech clipping for the purpose of SSB dynamic waveform reduction states (p 328):

> "Clipping may also be accomplished at SSB r-f, and this has the advantage that fewer in-band distortion products are created for a given amount of clipping.....Infinite clipping of the SSB signal followed by only enough filtering to remove the r-f harmonics would give a PEP-to-average power ratio of approximately 0 dB, since the result would approximate a constant-amplitude frequency-modulated r-f sine wave."

This process has become known as "RF clipping," and in a 1967 QST article[6], Sabin extolls RF clipping from the standpoint of producing

[2] Whalen, A. D., Detection of Signals in Noise, Academic Press, 1971, (pp 61-70).

[3] Bedrosian, E., "The Analytic Signal Representation of Modulated Waveforms," Proceedings of the IRE, October 1962, (pp 2071-2076).

[4] Dugundgi, J., "Envelopes and Pre-Envelopes of Real Waveforms," IRE Transactions on Information Theory, March 1953, (pp 53-57).

[5] Pappenfus, E. W., et.al., Single Sideband Principles and Circuits, McGraw-Hill, 1964.

[6] Sabin, W., "R. F. Clippers For S. S. B," QST, July 1967.

an essentially constant RF envelope with very low in-band distortion products.

Around 1975, the Principal Author/Investigator (J. C. Springett) was discussing the concept of RF clipping with a radio amateur who was having some difficulty implementing the technique, when it occurred to Springett that there should be a baseband equivalent to the RF clipping process. That is, there must exist a transformation of the baseband modulating signal which when input to an SSB modulator will produce a constant envelope RF signal. A summary of the analysis which establishes this equivalency follows.

Figure 3 is a block diagram of the functional RF clipper. When the baseband input to the SSB modulator is v(t), the SSB RF signal is

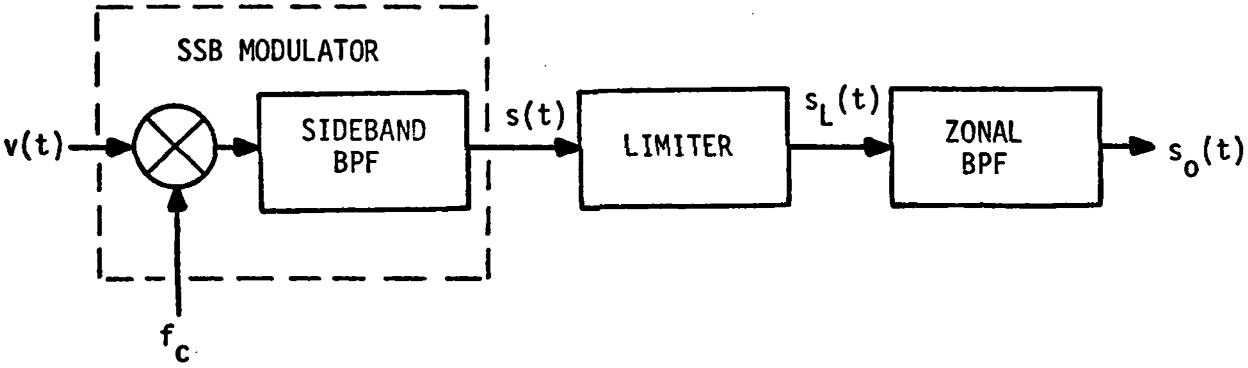


FIGURE 3 - RF CLIPPING CONFIGURATION

$$s(t) = v(t)\cos\omega_c t - \hat{v}(t)\sin\omega_c t, \qquad (5)$$

where $\hat{v}(t)$ is the Hilbert transform of $v(t)$. Alternatively, (5) may be written in the form

$$s(t) = a(t)\cos[\omega_c t + \theta(t)], \tag{6}$$

with

$$a(t) = \sqrt{v^2(t) + \hat{v}^2(t)}, \tag{7}$$

and

$$\theta(t) = \tan^{-1}\left[\hat{v}(t)/v(t)\right]. \tag{8}$$

The limiter (ideal signum function) removes the amplitude modulation, $a(t)$, and the limiter output becomes

$$s_L(t) = \text{Cos}\,[\omega_c t + \theta(t)]\ , \tag{9}$$

where $\text{Cos}(x)$ is given by[7]

$$\text{Cos}(x) = \frac{4}{\pi} \sum_{n=o}^{\infty} \frac{(-1)^n}{2n+1} \cos\,(2n+1)x\ . \tag{10}$$

Following the limiter, the BPF passes only the first zone of the limiter output, i.e, the $n=o$ term of (10), with the result (the $4/\pi$ coefficient is dropped)

$$s_o(t) = \cos[\omega_c t + \theta(t)], \tag{11}$$

which when expanded using (8) becomes

[7] Springett, J. C., and M. K. Simon, "An Analysis of the Phase Coherent-Incoherent Output of the Bandpass Limiter," IEEE Transactions on Communication Technology, February 1971.

$$s_o(t) = \frac{v(t)}{\sqrt{v^2(t) + \hat{v}^2(t)}} \cos\omega_c t - \frac{\hat{v}(t)}{\sqrt{v^2(t) + \hat{v}^2(t)}} \sin\omega_c t. \qquad (12)$$

This is in the form of an SSB signal (vis., eqn (5)) which means the equivalent baseband modulating signal form is

$$v_o(t) = \frac{v(t)}{\sqrt{v^2(t) + \hat{v}^2(t)}}. \qquad (13)$$

Equation (13), then, is the baseband signal equivalent to RF clipping.

3.2 Envelope Normalization Fundamentals

The operation defined by (13) is known as envelope normalization, wherein the speech signal is divided by its own true envelope to produce a constant envelope waveform $v_o(t)$. The true speech envelope is given by (as per Dugundji[4])

$$e(t) = \sqrt{v^2(t) + \hat{v}^2(t)}. \qquad (14)$$

Thus, EN is simply expressed as

$$v_o(t) = v(t)/e(t). \qquad (15)$$

Alternatively (15) may be written as

$$v(t) = e(t)v_o(t), \qquad (16)$$

which, when compared with (1) establishes that $g(t) = v(t)/e(t)$.

The formation of the true speech envelope is the key to the EN process, and in turn involves the calculation of the Hilbert transform (HT) of v(t), viz.,

$$v(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{v(\lambda)}{\lambda - t} d\lambda \qquad (17)$$

Clearly, exact determination of the HT involves the complete temporal record of v(t). In reality, this is not practical. Therefore, the Hilbert transform integral must be approximated by integrating between a finite set of limits, 0 to T. In terms of mechanizations, a finite T-second segment of v(t) must therefore be committed to some form of memory, while the speech signal itself must be delayed by T/2 seconds. That the equivalence of memory is manifest in the RF clipping configuration of Figure 3 should be understood from the fact that production of a high quality SSB signal, using the balanced-modulator/filter method, requires a multipole sideband BPF which inherently introduces transfer delay. Practical methods for calculating the finite Hilbert transform are discussed in subsections 5.2 and 5.3.

Another way of viewing the HT is in the transfer function sense. Inspection of (17) shows that the integral represents a convolution between the speech signal v(t), and the function $1/(\pi t)$ which is the impulse response of a network that produces an HT at its output. The transfer function of this network is

$$H(\omega) = j\mathrm{sgn}(\omega), \qquad (18)$$

where sgn ( ) is the signum function. The magnitude and phase functions are respectively:

$$|H(\omega)| = \begin{cases} 1 & \omega \neq 0 \\ 0 & \omega = 0 \end{cases} , \qquad (19)$$

$$\phi(\omega) = - \frac{\pi}{2}\mathrm{sgn}(\omega). \qquad (20)$$

It is seen, therefore that the HT is an all-pass network (except for ω=0) that shifts the phase of all frequency components of the input waveform by $-\pi/2$ or -90^{o} ($\omega>0$). When the limits on the HT integral become finite, then the magnitude and phase of the equivalent network become:

$$|H(\omega)| = \frac{2}{\pi}\,\mathrm{Si}(|\omega|T/2), \tag{21}$$

$$\phi(\omega) = -\,\frac{\pi}{2}\,\mathrm{sgn}(\omega), \tag{22}$$

where Si( ) is the sine-integral.

The phase shift property remains exact, while the magnitude is no longer all-pass. Figure 4 illustrates the exact and finite HT network magnitude and phase functions. Mechanizations of networks which approximate the HT are discussed under subsection 5.1.

Circuits capable of performing the complete envelope normalization operation are examined in subsection 6.1.

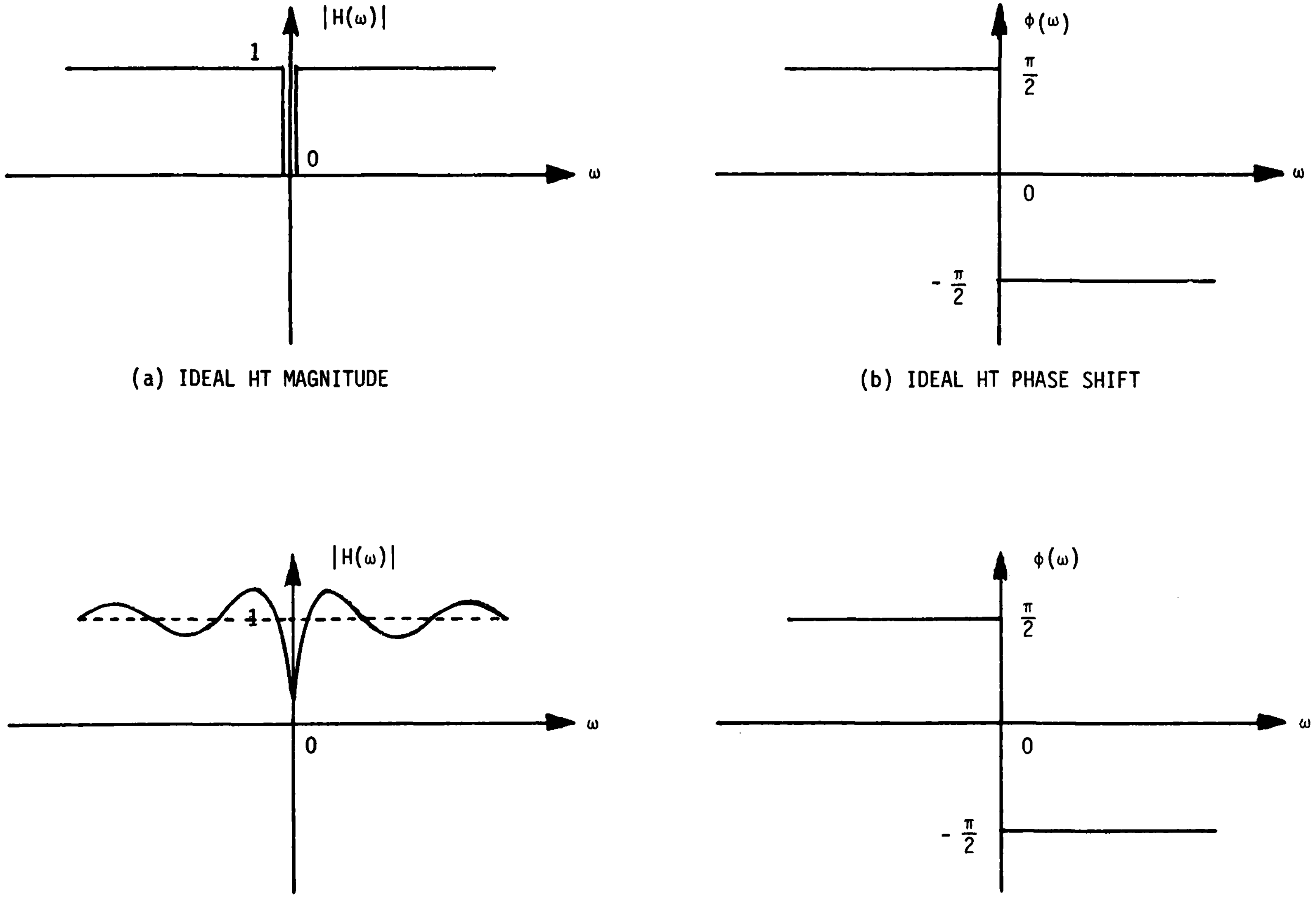
3.3 Envelope Normalization Properties

The following review of EN properties assumes that the HT is ideal. When the HT is approximate but of high quality, the properties from all practical standpoints remain valid.

First, recourse to eqns. (11) through (15) establish that

$$v(t)/e(t) = \cos\theta(t), \tag{22}$$

i.e., the envelope normalized speech may be viewed as a sinusoidal function of the instantaneous phase process $\theta(t)$ given by (8). Further, from (11) and (12) it appears that $\cos\theta(t)$ and $\sin\theta(t)$ are HT pairs because (12) is reported to be a SSB signal. (Note: this has not been rigorously proved for arbitrary $v(t)$ functions, but has been established for periodic and deterministic signal examples.)



(a) IDEAL HT MAGNITUDE

(b) IDEAL HT PHASE SHIFT

(c) FINITE HT MAGNITUDE

(d) FINITE HT PHASE SHIFT

FIGURE 4 - IDEAL AND FINITE HT CHARACTERISTICS

Thus, $\hat{v}(t)/e(t) = \sin\theta(t).$ (23)

The envelope, e(t), is always greater than or equal to zero. A set of inequalities which prove useful for the adjustment of one form of EN circuits is

$$v(t) + e(t) \geq 0, \text{ all } t, \tag{24}$$

and

$$\hat{v}(t) + e(t) \geq 0, \text{ all } t. \tag{25}$$

Using (22), (24) may easily be established, viz.,

$$v(t) + e(t) = [1 + \cos\theta(t)] \,.\, e(t), \tag{26}$$

and since e(t), ≥ 0 (by definition) and $1 + \cos\theta(t) \geq 0$, (24) results. Eqn. (25) is likewise proven using (23). The use of these inequalities is discussed in subsection 6 1.

The EN process alters the spectrum of the speech signal, v(t), and because EN involves nonlinear operations, it can be expected that out-of-band components will be generated by EN of a strictly bandlimited signal. The specific question to be addressed, then, is: if v(t) is a random process with non-constant envelope (e.g., speech), and spectrum $S_v(f)$, then what is the power spectrum $S_{v_o}(f)$ of $v_o(t) = v(t)/e(t)$? Although it would be desirable to compute $S_{v_o}(f)$ given only the power spectrum $S_v(f)$, it has been concluded that, in general, such is not possible because evaluation of the second order statistics (e.g., correlation function or power spectral density) of the result of passing a signal through a zero memory nonlinearity (such as the EN process) depends on the input process statistics of <u>all</u> orders.

Note that for a linear operation on a signal that is not true, and indeed the second order statistics of the output process depend only on the second order statistics of the input.

In order, then, to analytically study the spectrum changes resulting from EN, some form of deterministic or periodic signal must be used. However, in order for the results to be useful as a model for what will happen to real speech signals, the signal must have a slowly varying envelope with respect to the signal frequency components, i.e., the principal frequencies which comprise the envelope should be less than the lowest frequency component of the signal. This restriction severely limits the analytical choices. For example, the deterministic signal sin(t)/t does not meet the criterion, although the HT and Fourier transforms are easily calculated. The signal $\sin(at)J_0(bt)$ does satisfy the restriction provided $b<<a$, the HT is straightforward, but the Fourier transforms are not available in closed form.

Periodic functions produce the only useful analytical results because the input and EN signals are expressed as Fourier series. Thus, a finite (bandlimited) Fourier series may be specified, and the Fourier coefficients for the EN waveform then calculated. The principal problem now is designating functions that are amenable to deriving the envelope function in closed form.

One way of satisfying the slowly varying envelope restriction is to specify a low frequency, finite term, periodic function, m(t), and modulate it onto a "carrier" of frequency f_0. The result is

$$x(t) = m(t)\cos\omega_0 t, \tag{27}$$

where m(t) is lowpass and contains only frequencies that are less than f_0. Use of the HT product theorem gives

$$\hat{x}(t) = m(t)\sin\omega_0 t, \tag{28}$$

and the envelope becomes

$$e(t) = |m(t)| \tag{29}$$

Thus, the EN signal is

$$x_o(t) = \mathrm{sgn}\{m(t)\}\cos\omega_o t. \tag{30}$$

This form makes the EN Fourier series relatively easy to calculate given the Fourier series of m(t).

Two symmetrical input spectra are postulated as examples from which general conclusions may be inferred. The first spectrum is uniform in amplitude, and the input signal is given by the finite Fourier series.

$$x(t) = \sum_{k=\pm 1,\pm 3,..}^{\pm(N+1)} \sin(\omega_0 t + k\Delta\omega/2), \tag{31}$$

where N, the number of sinusoids spaced by $\Delta\omega$, is odd for the case at hand. Following EN, the output signal has the Fourier series

$$x_o(t) = \sum_{k=\pm 1,\pm 3,..}^{\pm\infty} \beta_k \sin(\omega_0 t + k\Delta\omega/2), \tag{32}$$

with the amplitude coefficient β_k being

$$\beta_k = 2/(\pi k)\tan(\pi k/2N). \tag{33}$$

The summation limit in (32) of ±∞ shows that EN has spread the finite spectrum into an infinite spectrum. The envelope signal also has an infinite spectrum given by

$$e(t) = \varepsilon_0 + 2 \sum_{n=1}^{\infty} \varepsilon_n \cos(n\Delta\omega t), \tag{34}$$

with

$$\varepsilon_n = \frac{1}{N} \sum_{k=1}^{\frac{N}{2}+n} \frac{\tan(\mu_k)}{\mu_k} + \frac{1}{N} \sum_{k=1}^{\left|\frac{N}{2}-n\right|} \frac{\tan(\mu_k)}{\mu_k} \operatorname{sgn}\left(\frac{N}{2}-n\right), \tag{35}$$

and

$$\mu_k = (2k-1)\pi/(2N). \tag{36}$$

From (33) it is seen that spectral coefficients do not depend on $\Delta\omega$ but are a function of N. This is likewise true for the envelope Fourier series coefficients. Nor do the coefficients depend upon the center frequency f_0.

Figure 5 shows the input and output spectra for N = 4 in both linear (power) and log (dB) forms. The input and output spectra are normalized to have the same total power. From the linear output spectrum plot it is readily seen that the two central components have been attenuated, while the other pair of input components have been slightly amplified. Very importantly, most of the power lost from the two central components has become manifest in the pair of components that appear at the frequencies $\pm 3\Delta f/2$. In fact, the percentage of total power in the 4

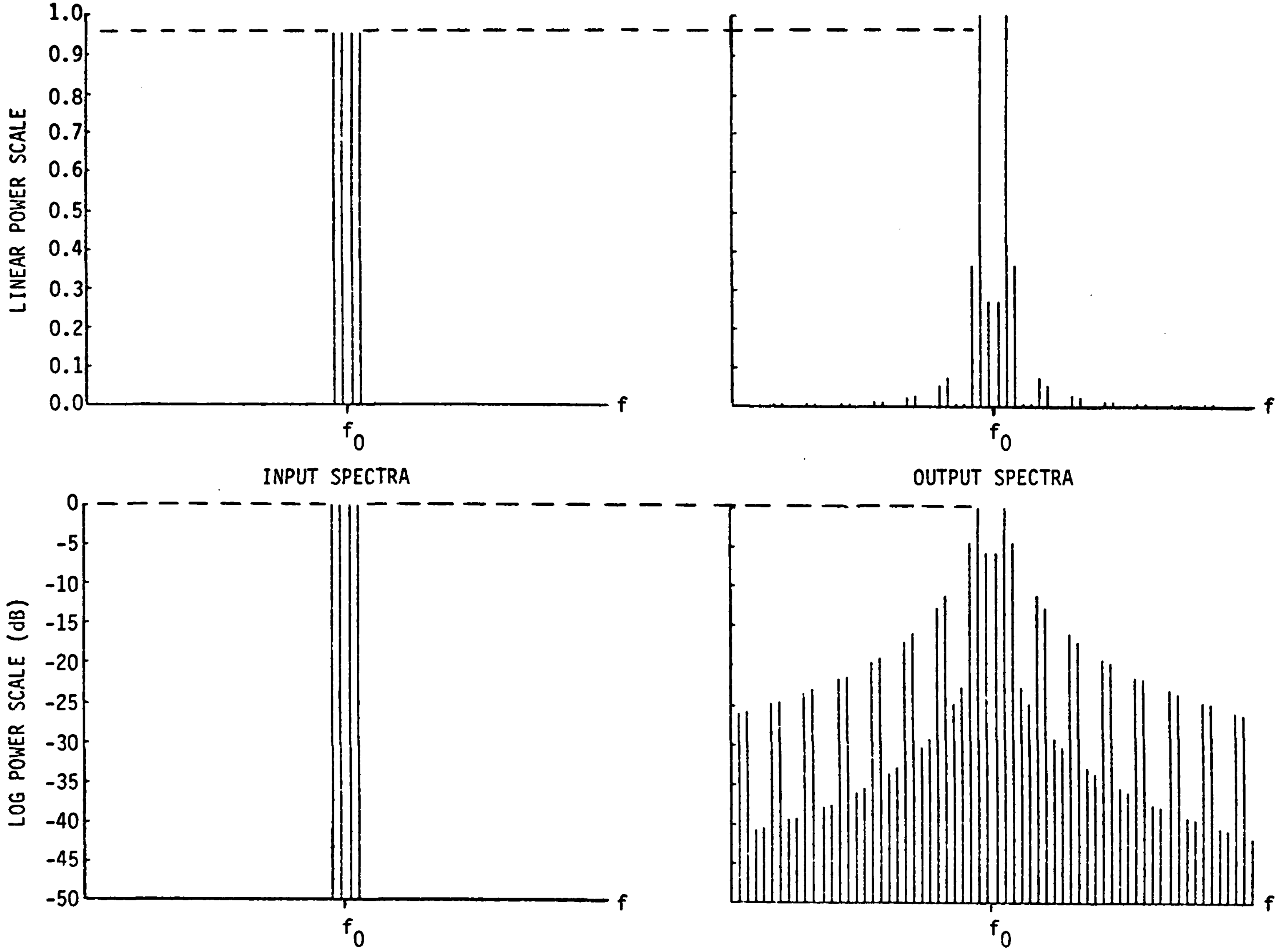


FIGURE 5 - EN INPUT AND OUTPUT SPECTRA FOR N = 4

central components (i.e., those that represent the input spectrum frequencies), is 66.4%, while the percent power in the two ±3$\Delta f/2$ components is 18.9%. Thus, the power in the six central components is 85.3% of the total power.

Figure 6 shows the corresponding spectra for N = 10. Again, most of the power lost from the original input spectrum is manifest in the two components that fall just outside of the input spectrum limits. In this case the percentage of total power represented by the first 12 central components is 80.6%.

The envelope spectra for N = 4 and N = 10 are plotted respectively as Figures 7 and 8. These spectra are seen to fall-off very rapidly from k = 0 (the zero frequency component), with 96% of the envelope power residing in the first three components for N = 4, and the first seven components when N = 10.

What is generally seen by these results is that the bandwidth expansion of the input signal due to the EN process is small, being confined to a few Δf above and below the bounds of the input spectrum. The larger the value of N, the lower the percentage of bandwidth expansion. Thus, the simple analytical model appears to confirm what has been observed in practice for complex speech signals (see subsection 9.2 for measured results).

A real speech spectrum begins to roll-off below and above some frequency where the spectrum is maximum (around 400 Hz for the composite speaking population[8]). The second input spectrum example therefore is one that is maximum at f_o and falls-off symmetrically to a level 10 dB below the maximum. Figure 9 shows the spectra plots for N = 10. The power lost from the ten input terms is virtually all manifest in the two components falling immediately outside of the input spectrum limits.

[8] Jakes, W. C., editor, Microwave Mobile Communications, John Wiley & Sons, 1974.

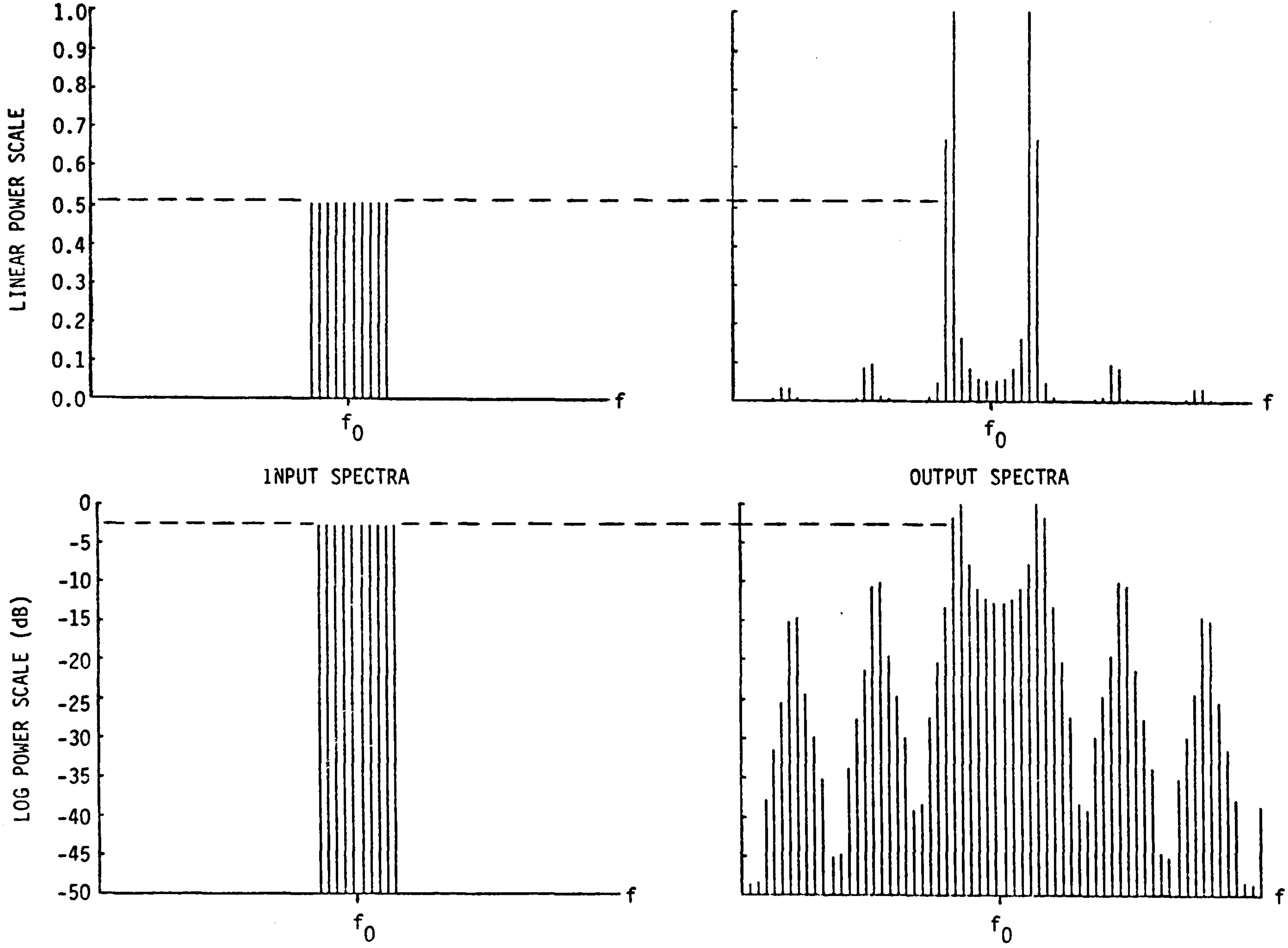


FIGURE 6 - EN INPUT AND OUTPUT SPECTRA FOR N = 10

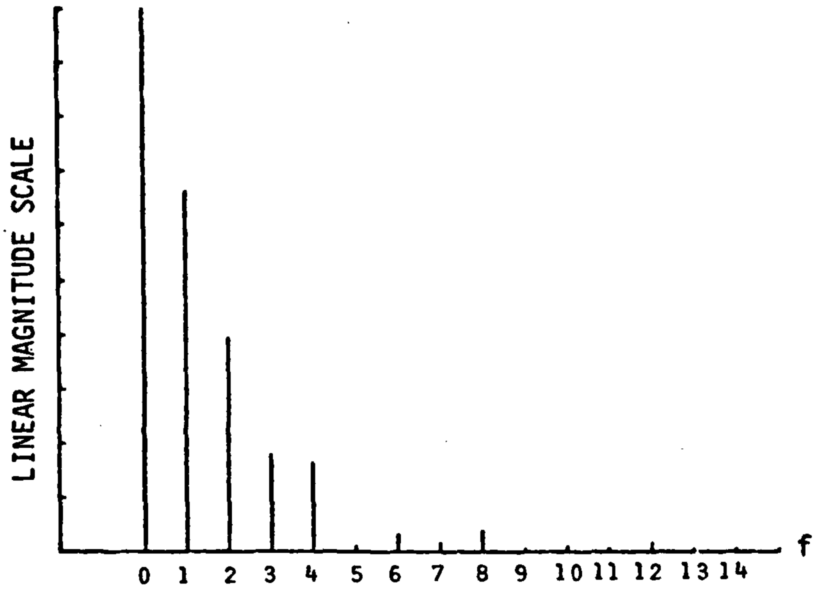


FIGURE 7 - ENVELOPE SPECTRUM, N = 4

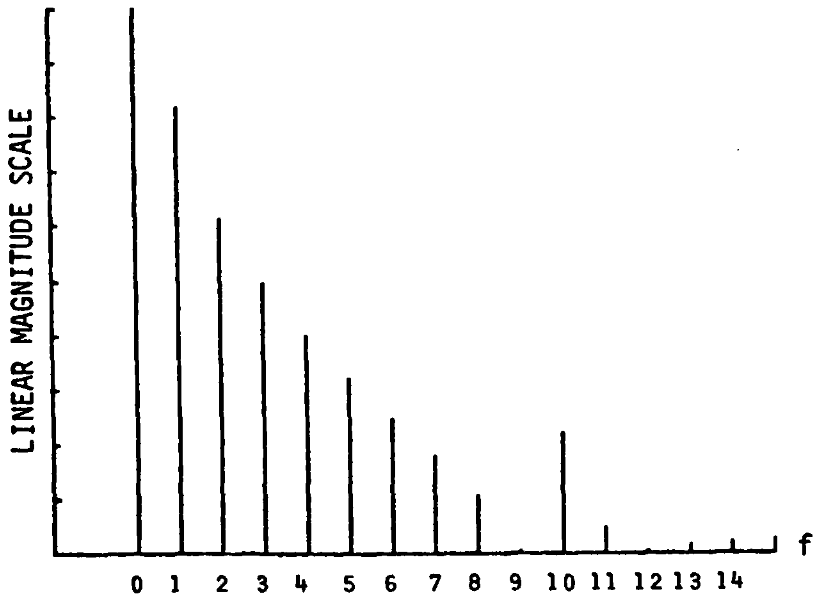


FIGURE 8 - ENVELOPE SPECTRUM, N = 10

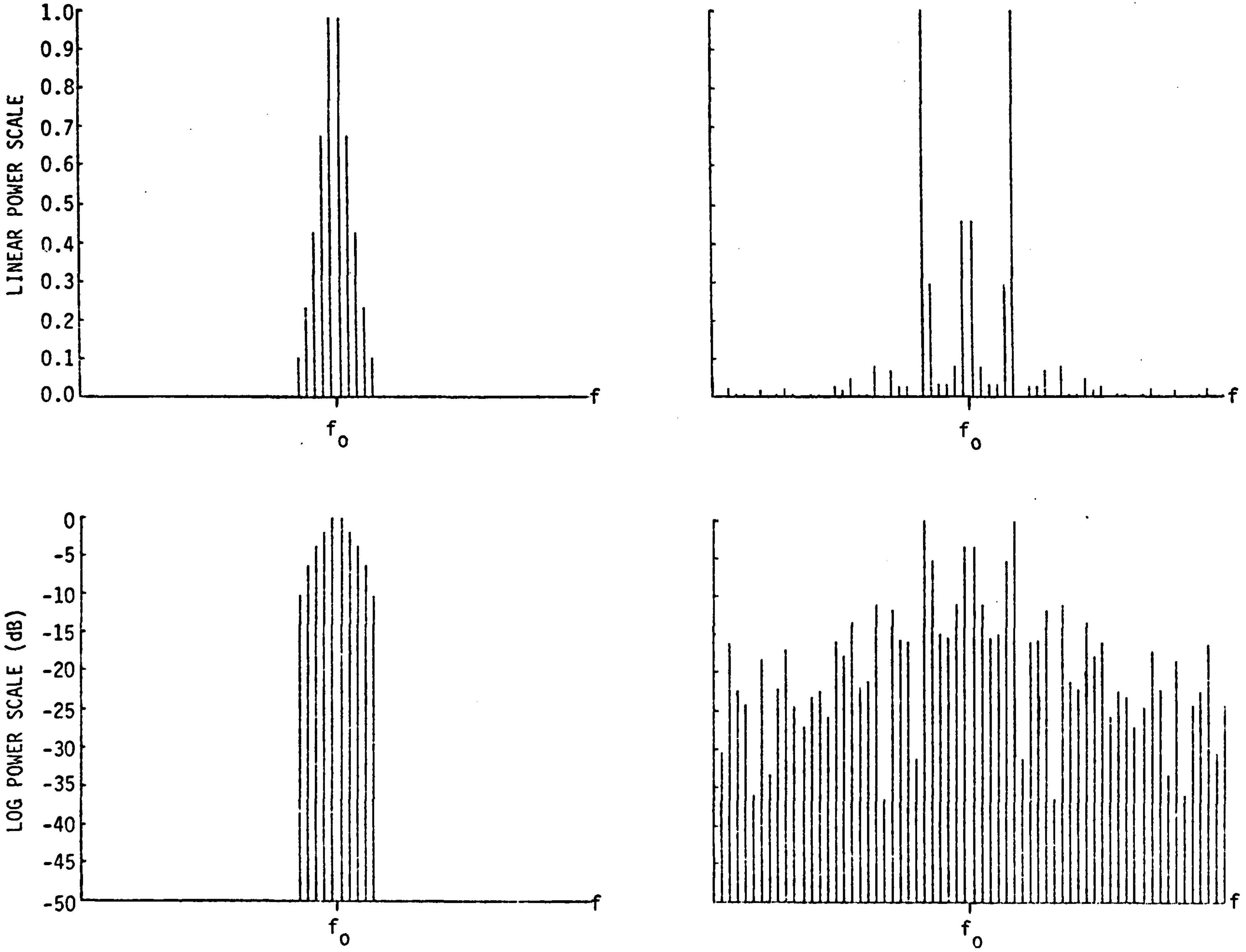


FIGURE 9 - EN TILTED INPUT AND OUTPUT SPECTRA FOR N = 10

The power within the central 12 components if 78.2% of the total power. It may be concluded that this new example of spectrum expansion due to EN does not alter the general conclusions made previously.

The final subject considered in this subsection is the effect of EN on signals where noise (acoustical or electrical) accompanies the speech. Let the input signal to the EN operation be

$$x(t) = v(t) + i(t), \tag{37}$$

where i(t) is any interfering signal, including extraneous speech. The envelope is

$$e_x(t) = \left\{ \left[v(t) + i(t)\right]^2 + \left[\hat{v}(t) + \hat{i}(t)\right] \right\}^{1/2}, \tag{38}$$

and the EN signal becomes

$$x_o(t) = v(t)/e_x(t) + i(t)/e_x(t) \tag{39}$$

EN will alter the speech to interference ratio (SIR), as does any nonlinear operation on desired signal plus interference. Calculation of the EN SIR depends upon the statistics of v(t) and i(t), and exact evaluation is beyond the scope of this report (because of the difficult, and somewhat undefined, statistics involved). Nevertheless, general expected results may be easily reasoned. During those temporal segments when the speech dominates the interference, the interference will be suppressed, improving the SIR. On the other hand, during speech pauses, however short, the interference will dominate, and for ideal EN will be amplified to the full EN output level. Subjectively, to a listener of EN speech, this effect could prove more detrimental than the speech present SIR improvement proves beneficial. As a result, EN speech may appear to sound noisier than the original. Of course, expansion of the EN signal at the receiver by multiplication with the envelope completely negates this property. However, there may be applications (e.g., when a large amount of acoustical noise is present at the listening point), wherein expansion might not be desirable (because it decreases the level or volume of the weaker syllables critical to intelligibility). Thus, during those periods when speech is absent, the output of the EN function should be blanked (switched to zero) to prevent the highly amplified interference from reaching the listener. Even when expansion is employed, blanking of the interference into the link transmitter provides benefits in terms of average transmitter power and adjacent channel interference considerations. The process by which blanking is accomplished involves speech vs. "silence" (i.e., noise only) discrimination/detection, and is often referred to an VOX (voice operated transmission or switch). VOX methods and their effectiveness are discussed in subsection 6.2.

4.0 APPLICATION OF ENVELOPE NORMALIZATION TO THE IMPROVEMENT OF COMMUNICATION SYSTEM PERFORMANCE

4.1 Direct SNR Improvement

As discussed in subsection 2.1, the communication link is usually designed to meet some minimum requirement for the soft spoken talker. Experience has shown that when the soft talker's weak syllable SNR is 12 dB, an overall 90% intelligibility is obtained and the subjective listening quality is judged to be "fair."[9] The presence of background noise is very noticeable. Such performance is some 8-10 dB below minimum wireline telephone quality, and is often referred to a communication or field quality.

For a constant level of noise at the receiver, speech compression at the transmitter has the effect of changing the received speech SNR as a function of speaker level and the compression ratio used. If an SNR of 12 dB is assigned to the soft spoken RMS level with no compression, then the SNR's tabulated in Table 2 will result (see Table 1 for basis.) The unaffected SNR is the RMS SNR of the average speaker. Thus, as is seen from Table 2, the EN SNR becomes 37 dB for all speakers at all speaking levels. With respect to the soft speaker weak syllable SNR, the SNR increase in seen to be 25 dB. Note that the loud speaker SNR's are correspondingly decreased.

The unaffected level or SNR may, of course, be set anywhere desired. For military applications, establishing the unaffected SNR as that of the RMS SNR for the average speaker, resulting in the EN speech SNR of 37 dB, may represent overdesign in terms of EN transmitter power or bandwidth. Setting the uneffected SNR to, say, the average speaker weak syllable SNR of 27 dB may prove more efficient. This still

[9] Lusignan, B.,"The Use of Amplitude Compandored SSB In The Mobile Radio Bands: A Progress Report," Stanford Radioscience Laboratory, Feb. 1980.

SPEAKING POPULATION	NO COMPRESSION	2:1 COMPRESSION	4:1 COMPRESSION	ENVELOPE NORMALIZATION
LOUD SPEAKER				
Strong Syllable SNR	62.0 dB	49.5 dB	43.25 dB	37.0 dB
RMS SNR	52.0 dB	44.5 dB	40.75 dB	37.0 dB
Weak Syllable SNR	42.0 dB	39.5 dB	38.25 dB	37.0 dB
AVERAGE SPEAKER				
Strong Syllable SNR	47.0 dB	42.0 dB	39.50 dB	37.0 dB
RMS SNR (unaffected SNR)	37.0 dB	37.0 dB	37.00 dB	37.0 dB
Weak Syllable SNR	27.0 dB	32.0 dB	34.50 dB	37.0 dB
SOFT SPEAKER				
Strong Syllable SNR	32.0 dB	34.5 dB	35.75 dB	37.0 dB
RMS SNR	22.0 dB	29.5 dB	33.25 dB	37.0 dB
Weak Syllable SNR	12.0 dB	24.5 dB	30.75 dB	37.0 dB

TABLE 2 - SPEECH SNR's AS A FUNCTION OF VOICE COMPRESSION RATIO

represent a soft speaker weak syllable SNR gain of 15 dB. The potential gain in either transmitter power and/or bandwidth is discussed under subsection 4.3.

4.2 Expandor Effects; Subjective SNR Improvement

Figure 10 illustrates the basic ideal expandor operation and parameters. The EN speech signal given by (15) is input to the multiplier along with the additive noise, n(t). Multiplication by the speech envelope, which is assumed ideally to be noise free (performance with a noisy speech envelope is discussed in subsection 8.2), yields

$$r_o(t) = v(t) + e(t)n(t). \tag{41}$$

The first property of expansion is that the real speech SNR is unchanged through the expandor. With the overbar designating the expected value, the following definitions are made:

$$\overline{v^2(t)} = \sigma_v^2 \tag{42}$$

$$\overline{e^2(t)} = 2\sigma_v^2 , \tag{43}$$

$$\overline{v_o^2(t)} = 1/2 , \tag{44}$$

$$\overline{n^2(t)} = \sigma_n^2. \tag{45}$$

Equation (43) follows directly from (14) using the fact that v(t) and $\hat{v}(t)$ are orthogonal and have equal power, and (44) is obtained using (22). The SNR into the expandor is

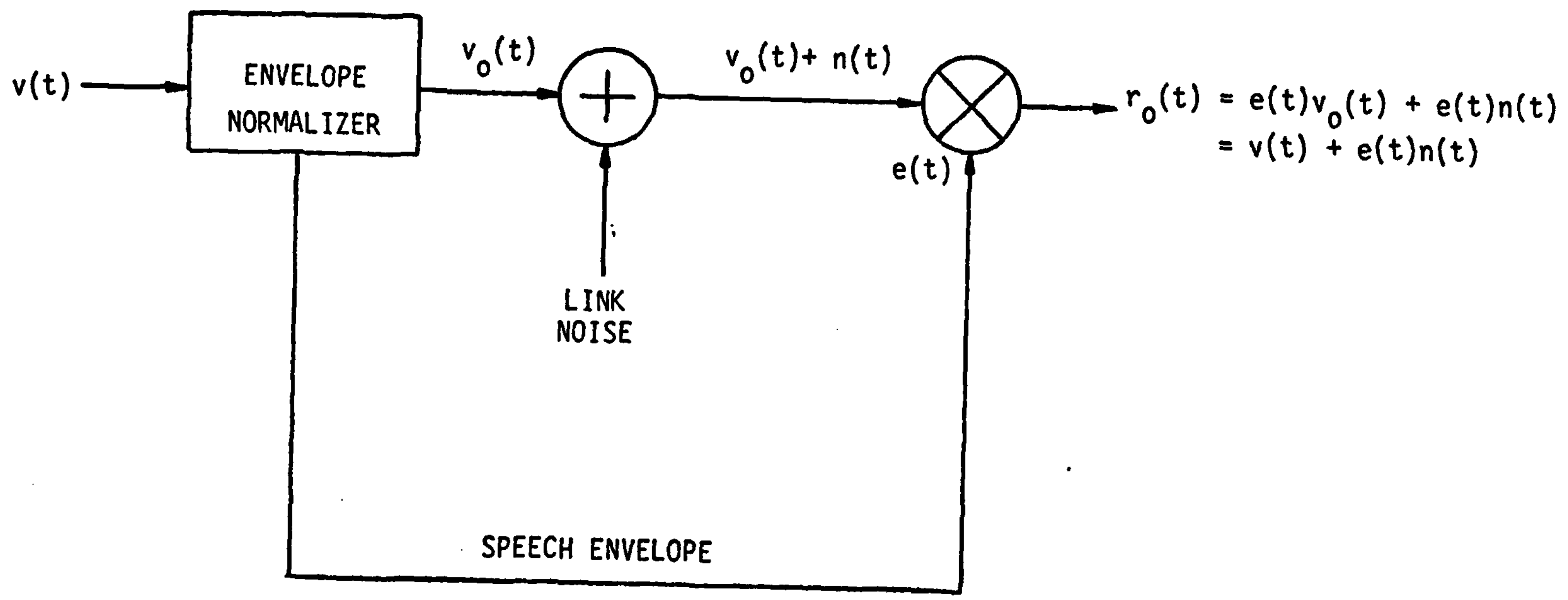


FIGURE 10 - IDEAL EXPANDOR OPERATION

$$SNR_1 = \overline{v_o^2(t)}/\overline{n^2(t)} = 1/2\sigma_n^2 , \qquad (46)$$

while the SNR at the expandor output becomes

$$SNR_2 = \overline{v^2(t)}/\overline{e^2(t)n^2(t)} = \sigma_v^2/(2\sigma_v^2\sigma_n^2) = 1/2\sigma_n^2 , \qquad (47)$$

thus proving the property.

The second aspect of expansion is that it results in what is known as a subjective SNR improvement. Referring to (41), because e(t) is zero during all speech pauses (however short), then the noise to the listener is also zero during these pauses. Thus, the listener will not hear noise when speech is absent, which results in a subjective improvement relative to a receiving system which does not employ expansion. In other words, the listener perceives that a receiver with an expandor is not as noisy as a receiver that does not have an expandor.

Since EN in the context of this Report has not appeared in the literature, no prior basis for EN expandor subjective SNR improvement therefore exists. All available references discuss subjective SNR improvement for 2:1 companding, such improvement being stated to range from 10 dB to 20 dB. These values are a function of speech SNR, being greatest for the higher SNR's (>30 dB). Virtually none of the references state the basis for subjective SNR improvement, nor do they explain how it is measured. Only one reference[10] gives some insight into the problem. Basically, the improvement consists of two components: (1) the increase in speech SNR (due to compression) when speech is present, and (2) the decrease in noise level (due to expansion) when speech is absent. Further, the overall subjective improvement is a

[10] Richards, "Transmission Performance Assessment For Telephone Network Planning," Proc. IEE, May, 1964.

weighted function of (1) and (2) (based upon undocumented listening tests,) being 1/3 of (1) plus 2/3 of (2). Additionally, the overall improvement is a function of talker level, being the greatest for a weak talker.

It should be noted that Richards' formula for subjective improvement involves the direct SNR enhancement discussed in subsection 4.1. Since this is an objective or measurable quantity, it really should not be included as part of the subjective SNR improvement. Subjective improvement should only reflect the apparent decrease in perceived noise level due to expansion.

Richards' formula for 2:1 companding is now briefly reviewed, it being given as

$$\Delta SNR = \frac{U - S}{6} + \frac{2(U - N)}{3} \tag{48}$$

where

S = mean power of speech into the compressor (dBx),
N = mean audio noise power (dBx),
U = unaffected companding level (dBx),

with dBx representing the quantities in decibels relative to any reference x. Taking an example of a soft speaker whose uncompanded RMS SNR is 22 dB, and using Tables 1 and 2, the parameters may be specified as

$S = -15$ dB
$N = -37$ dB
$U = 0$ dB,

the total improvement becomes $\Delta SNR = 2.5 + 24.7 = 27.2$ dB. This appears to be a very large and perhaps unrealistic improvement. Notice that the effective improvement due to compression is but 2.5 dB compared to 7.5 dB actually obtained (see Table 2). Most of the subjective improvement occurs due to quieting during the speech pauses. Further, the result is dependent on the unaffected level. If this level were to be changed from the average speaker's RMS to weak syllable level (i.e., U = -10 dB), then the total improvement would become $\Delta SNR = 0.8 + 18 = 18.8$ dB. This result is more in line with the range of expected improvement reported in the literature, and is consistent with the practice of setting the unaffected level of the compandor below the median of the speaking population.

When Richards' basic definition of subjective improvement is applied to EN, it is found that the improvement becomes infinite because the decrease in the noise level due to expansion when speech is absent is total, i.e., there is absolutely no noise output from the expandor. This result, of course, is totally unrealistic. All that can be concluded is that Richards' approach is heuristic, and is applicable only to the 2:1 companding law. What, then, can be conjectured concerning the subjective SNR improvement when using EN? First, the direct SNR improvement due to compression is tangible and greater than that obtained with 2:1 compression. Secondly, the perceived decrease in receiver noise due to expansion will be more with EN than with 2:1 companding, but not extensively more. No further improvement is obtained when the noise is attenuated below the threshold of hearing, or significantly below the level of other listening disturbances. Suffice it to say that if, according to Richards' concept for subjective SNR improvement, a 10 dB to 20 dB range of increase should be expected with 2:1 companding, then using EN a comparable range of 15 dB to 30 dB may be presumed. Experimental verification of this, however, is very difficult, and will be discussed further in subsection 9.4.

4.3 Advantages of EN When Applied to Radio Systems

In subsection 4.1, the direct SNR improvement due to EN compression was examined apart from any consideration of the radio link itself. In other words, the gains obtained were with respect to a simple linear, additive noise, channel. Because AM and SSB are akin to such a channel under certain conditions, the expected improvements may be directly realized. With FM, the effective gains require additional consideration. The advantages of EN when applied to SSB and FM radio systems are discussed in the following paragraphs.

Equation (5) represents a SSB signal where the modulation is unprocessed speech, while (12) is the SSB signal with EN. An SSB transmitter is usually rated in terms of its peak envelope power (PEP) which for unprocessed voice is typically 17 dB above the power output for the average speaker's RMS level. It is assumed that some form of ALC is employed, otherwise the PEP rating would have to be approximately 28 dB above the average speaker's RMS level in order to accommodate a loud speaker without generating distortion. Referring to (11), (12), (43), and Table 1, the peak envelope power required for EN is found to be 3 dB above any speaker's RMS level (after EN compression). Thus, equating PEP in both cases gives a 14 dB advantage to EN in terms of direct speech SNR improvement (as the receiver, consisting of a product detector, is linear and therefore the speech SNR is directly proportional to the transmitter's RMS power). Of course, a transmitter having a PEP rating of 17 dB may not be able to accommodate an EN-SSB signal at the maximum level without overheating or exceeding the average power available from the power supply, in which case the 14 dB advantage would have to be reduced to a value commensurate with the maximum operating capabilities of the transmitter.

Conventional SSB transmitters employ class-A power amplifiers in order to accommodate the dynamic range of the speech modulated RF envelope. However, with EN, the RF envelope is constant (eqn. (11)), and as a result, for new SSB system designs which make use of EN only, a much higher efficiency class-C power amplifier may be employed. Note also, because of the constant envelope nature of EN-SSB, that nonlinear repeaters may be used to expand the coverage range of the application (as is done with mobile FM, etc.).

FM reception generally depends upon two parameters, the transmitter power and carrier frequency deviation. The received speech SNR is given by

$$\mathrm{SNR} = \frac{3P\,D_f^2\,\overline{m^2(t)}}{N_o\,f_m^3} \tag{49}$$

where,

> P is proportional to the transmitter power,
> D_f is the modulation sensitivity (Hz/volt),
> m(t) is the modulation (speech),
> N_o is the receiver noise spectral density (Watts/Hz),
> f_m is the speech lowpass bandwidth (Hz).

If $\sqrt{\overline{m^2(t)}} = \sigma_m$, then $D_f\sigma_m = \Delta f$ is the RMS deviation of the transmitter (in Hz). Further, if the peak value of the modulation is $\delta\sigma_m$, then $D_f\lambda\sigma_m = \Delta f_p = \delta\Delta f$ is the peak deviation of the transmitter. For uncompressed speech $\delta = 3.5$, while for EN speech, $\delta = \sqrt{2}$.

Most FM radio systems are designed to accommodate a specified peak deviation. Thus, for the same peak deviation, EN speech will gain a $(3.5/\sqrt{2})^2$ (= 7.9 dB) SNR advantage over unprocessed speech (all other parameters equal). This result assumes the uncompressed speech is scaled to optimally deviate the transmitter in accord with the peak limitation. Generally, ALC may be assumed as the means for accomplishing such regulation, but may fail to achieve full scale deviation by a factor of two for the weakest of speakers using typical ALC circuits. Therefore, the EN advantage could be as large as 14 dB for such conditions. When no ALC is employed at the FM transmitter with uncompressed speech, the EN gain for a soft speaker could be as much as 48 dB!

For some applications, it may not be necessary to increase the speech SNR by employing EN. Rather, savings in FM transmitter power and/or channel bandwidth may be sought. Again, for the $3.5/\sqrt{2}$ deviation ratio, the transmitter power may be decreased by a factor of 6.1 (and even more for less optimum unprocessed speech conditions mentioned above). FM bandwidth is conveniently (and quite accurately for speech) given by Carson's rule, viz,

$$B = 2\ (\Delta f_p + f_m). \tag{50}$$

Letting B, correspond to $\Delta f_p = 3.5\ \Delta f$, and B_2 be the bandwidth for $\Delta f_p = \sqrt{2}\Delta f$, B_2 may be express in terms of B, as

$$B_2 = (\sqrt{2}/3.5)B_1 = 0.4B_1. \tag{51}$$

This result states that 2.5 EN-FM channels may be obtained for the bandwidth needed by uncompressed speech FM. Practically, the factor becomes 3 EN-FM channels when the previously discussed suboptimum conditions are considered. Typically B_1 = 30 KHz, so B_2 may be 10 KHz.

It should also be recalled that, even though transmitter power or channel bandwidth may be reduced by equating the EN speech and uncompressed speech SNRs, the EN system will still achieve the subjective SNR improvement due to exapndor quieting discussed in subsection 4.2. However, for this to become possible, the speech envelope is required at the receiver, and can be obtained only by transmitting it along with the EN speech itself. This process, known as linking, will modify, somewhat, the performance improvements discussed in this section. Envelope linking is examined in Section 8.0.

5.0 CIRCUITS FOR HILBERT TRANSFORM DERIVATION

In the following subsections, the term Hilbert transform (HT) will be applied in the generic sense to any circuit configuration that produces an output signal pair whose properties are akin to those described in subsection 3.2.

5.1 Wideband 90° Phase-Shifters

Studies of all-pass networks capable of approximating the HT over the audio frequency range (up to two decades), as per equations (19) and (20), have produced many in-depth results.[11,12,13] The early implementations used passive components (principally resistors, capacitors, and transformers), and were difficult and expensive to construct because the component values required an accuracy on the order of 1% or less (necessitating component synthesis by selection and trimming utilizing an impedance bridge). Several such networks have been constructed and tested by the author in past years, with their performance being judged only moderately satisfactory.

Clearly, for the EN applications envisioned, the preceding approach is not workable. The need for component high precision stems from the fact that the networks embody minimum pole/zero configurations for stated

[11] Weaver, D. K. Jr., "Design of RC Wide-Band 90-Degree Phase-Difference Network," Proceedings of the IRE, April 1954.

[12] Bedrosian, S. D., "Normalized Design of 90° Phase-Differenced Networks," IRE Transactions on Circuit Theory, June 1960.

[13] Tsuchiya, T., and S. Shida, "On the Design of Broad-Band 90° Phase-Splitting Networks," IEEE Transactions on Circuits and Systems, January, 1980.

performance specifications. A method for circumventing this problem is to use a higher order network than strictly required (i.e., theoretically exceed the specification), and then employ lower tolerance components so that the degenerative performance (due to inexact pole/zero locations) meets the desired specification.

A differential lead/lag circuit[14] based on this approach was constructed and evaluated. Figure 11 shows the configuration. Because a sixth-order network is employed, 10% tolerance components may be used without seriously compromising phase-shift performance. In fact, quadrature output is maintained to within ±3° over a frequency range of 300 Hz to 4 KHz. However, a 3 dB gain variation occurs over this frequency range. Since accurate signal envelope estimation requires both good phase and amplitude accuracy from the HT circuit, this was judged to be less than acceptable performance.

Note that the phase-shift network shown in Figure 11 is passive, although active input and output coupling is employed. A second approach[15] using a series of active sections is shown in Figure 12a. Each active section has the all-pass transfer function

$$H_i(\omega) = \frac{s + \alpha_i}{s - \alpha_i} \quad , \tag{52}$$

where $\alpha_i = 1/(R_iC_i)$. The phase shift through each section is given by

[14] DeMaw, D., editor, The Radio Amateur's Handbook, 1980 Fifty-Seventh Edition, American Relay Radio League, 1979.

[15] Williams, A. B., Electronic Filter Design Handbook, McGraw-Hill Book Company, 1980.

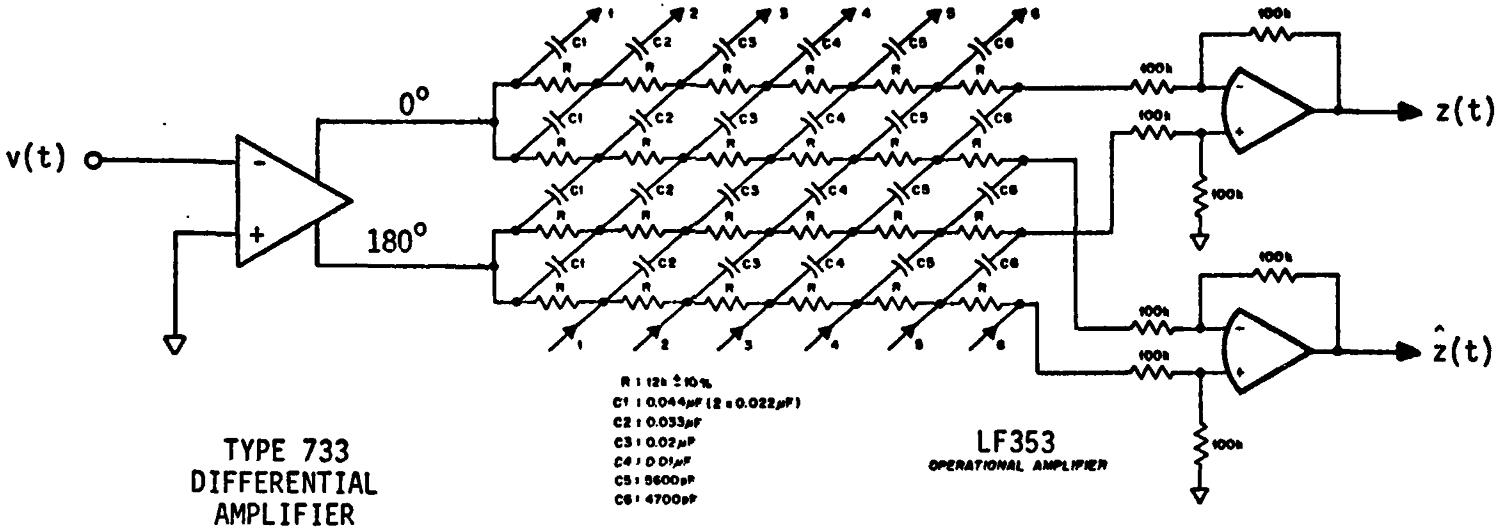
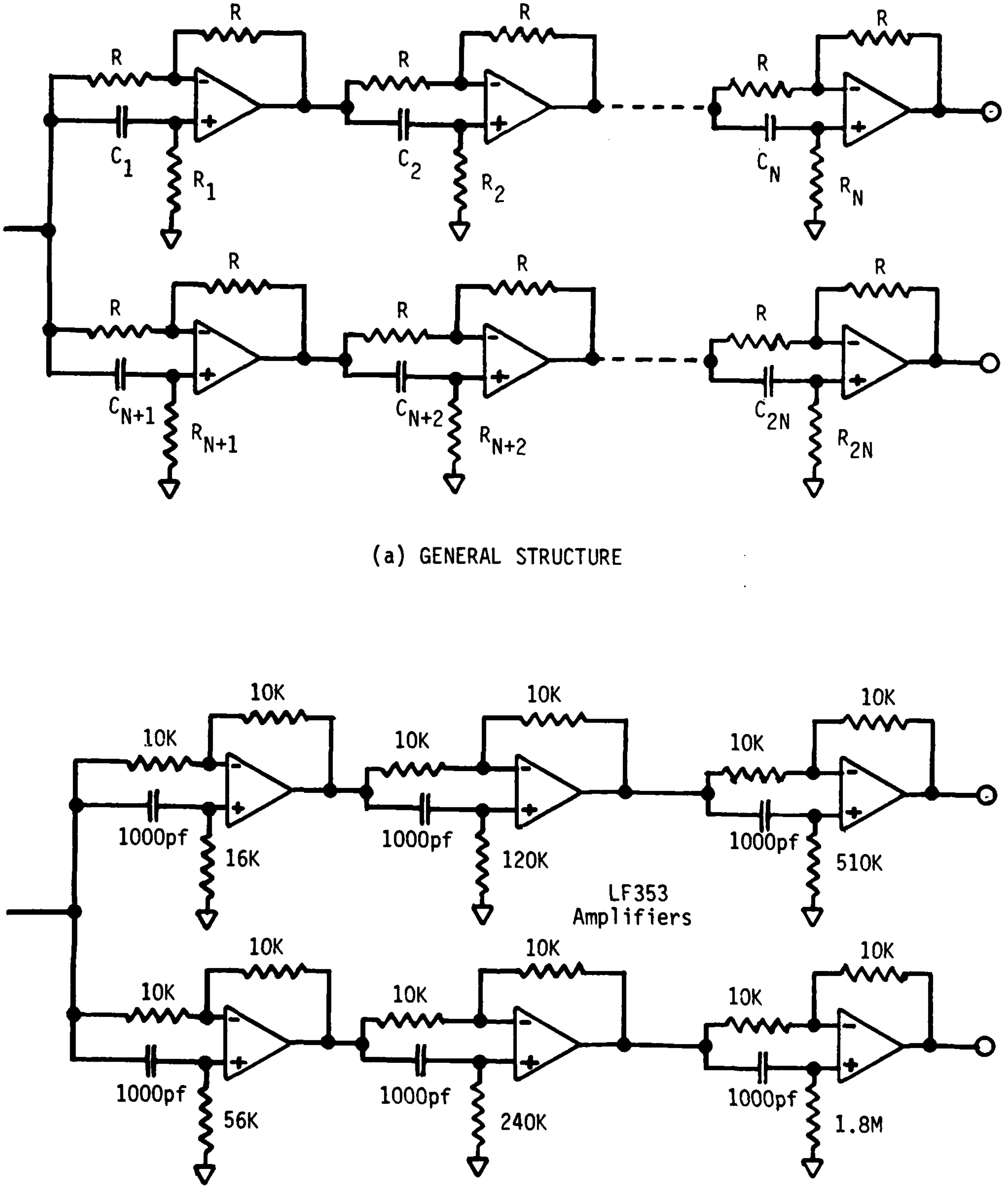


FIGURE 11 - DIFFERENTIAL LEAD/LAG 90° PHASE SHIFTER



(a) GENERAL STRUCTURE

(b) SPECIFIC CIRCUIT

FIGURE 12 - ACTIVE WIDEBAND 90° PHASE SHIFTER CIRCUIT

$$\phi(\omega) = -\tan^{-1}\left\{\frac{2\alpha_i\omega}{\alpha_i^2 - \omega^2}\right\}. \qquad (53)$$

It should be realized that as ω increases from $\omega = 0$ to $\omega \to \infty$, $\phi(\omega)$ goes from 0° to -180°.

To implement a wideband phase shifter, one leg is designed to provide 90° more phase shift than the other leg. (Note that the absolute phase shift through each leg is much greater than 90°.) Figure 12b shows a 6-section network, implemented with LF353 operational amplifiers, 5% tolerance resistors, and 20% tolerance capacitors. Measured results on this circuit have shown the differential phase output to be 90° ± 0.5°, and an amplitude variation of ± 0.3 dB, over a frequency range of 250 Hz to 3700 Hz. This performance is definitely superior to that obtained from the passive network. A parts count shows the active approach to also have a slight implementation edge over the passive network with its required input and output amplifiers, as indicated by the following tabulation.

	Active Network	Passive Network
Amplifiers	6	3
Resistors	18	24
Capacitors	6	24

This active network has been incorporated into the Envelope Normalization Demonstrator.

Probably the only disadvantage of the phase shift approach is that the exact shape of the input waveform, v(t), is not preserved at either of the output terminals. Thus, rather than obtaining v(t) and $\hat{v}(t)$, the outputs consist of z(t) and $\hat{z}(t)$, where z(t) is a waveform obtained by shifting all of the v(t) frequency components by whatever the wideband phase shift attributable to the upper leg of the network. As a consequence, the normalization process will result in z(t)/env {z(t)}.

For speech the result is perfectly acceptable because the ear is quite insensitive to this type of phase shift distortion. On the other hand, for non-speech types of baseband waveforms such distortion usually cannot be tolerated. As an example, suppose a digital (two-level) waveform is to be accommodated. Now, even though this waveform already has a constant envelope and therefore does not require envelope normalization, passing it through the EN process should not distort the waveform, which it will do if a broadband phase shifter is employed. (The direct type of delay line Hilbert transform discussed in subsection 5.2 does not suffer from this effect.) Thus, use of the wideband phase shift network approach may only be applicable to speech signals.

5.2 Sampled Data Direct Hilbert Transforms

The HT defined by (17) may be approximated in sampled data discrete-time form. As stated in subsection 3.2, the HT may be viewed as the convolution of the input signal with the HT impulse response. In discrete form, the input signal is sampled at its Nyquist-rate, a finite segment is stored in a memory or shift-register, and the convolution is carried out by forming a sum on the signal samples weighted by samples of the HT impulse response. The theory of the sampled data HT is given in Reference 16. Figure 13 is a functional block diagram of the discrete finite impulse response (FIR) HT based on a shift-register type of storage. The total number of stages is 2N, where N is an odd number. HT Theory establishes that the weights for samples appearing at the odd numbered stage outputs are zero. The non-zero weights have the property

$$W_{N+k} = -W_{N-k} , \tag{54}$$

for k = 1, 3, 5, ...

[16] Rabiner, L. R., and B. Gold, Theory and Application of Digital Signal Processing, Prentice-Hall, 1975.

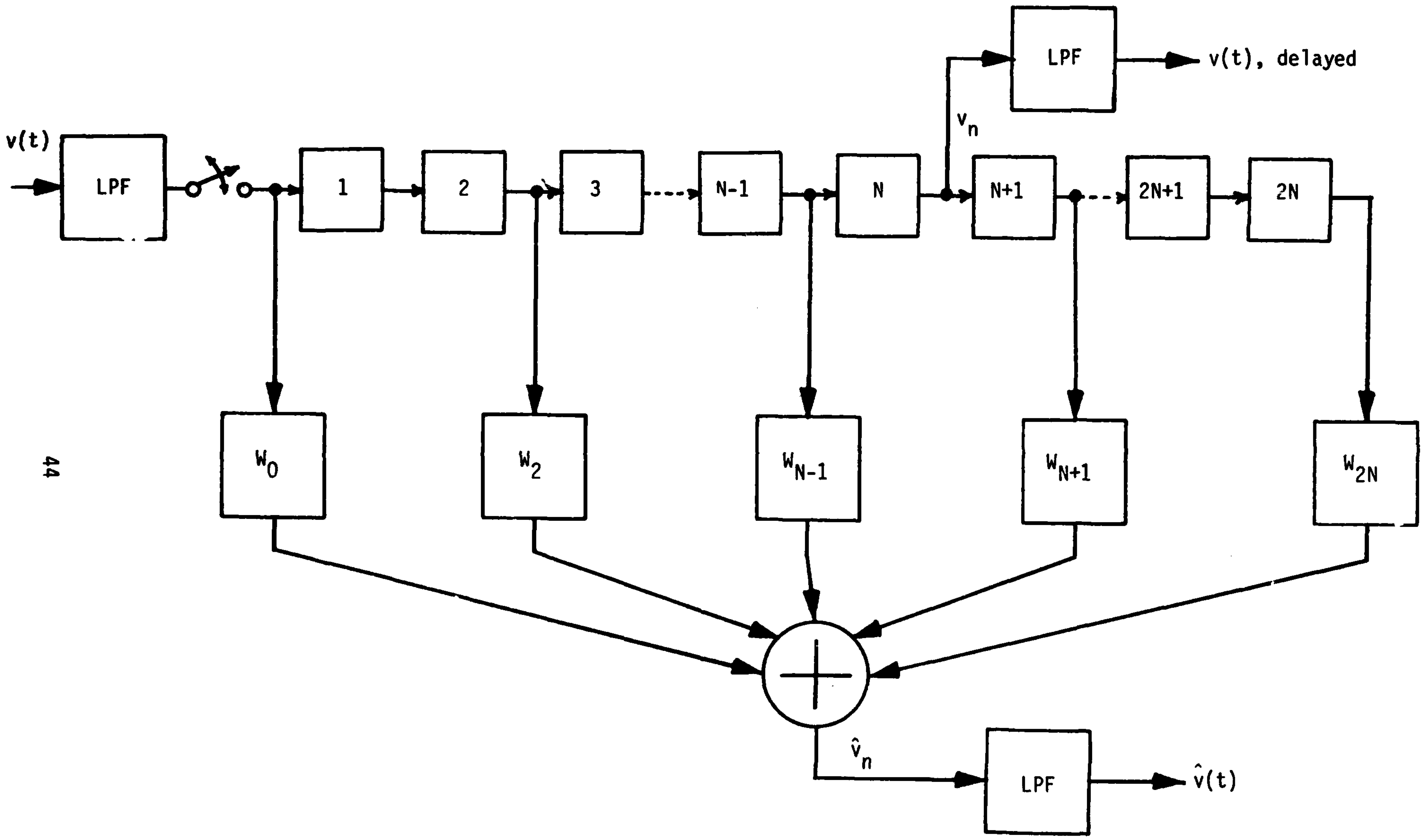


FIGURE 13 - SAMPLED DATA FIR HT CONFIGURATION

For a straightforward realization, the weights are taken as

$$W_{N+k} = 1/k. \tag{55}$$

However, this results in a sampled data transfer function magnitude, $|H(\omega)|$, akin to that of eqn. (21) and illustrated in Figure 4C. Since this response significantly departs from the ideal all-pass characteristic for low frequencies, it is less than desirable for application to speech envelope derivation (because the speech spectrum is maximum on the lower audio frequency range).

It is possible to modify $|H(\omega)|$, without affecting the $\pi/2$ phase shift, in such a way that the error relative to the ideal $|H(\omega)| = 1$ is minimized. Optimization in the minimax sense[17] alters the weightings somewhat from (55), but (54) still holds. The price paid is that the HT minimum error must be specified between fixed frequency limits, while outside of these limits the magnitude characteristic deviates markedly from ideal behavior (implying that the input signal should possess insignificant components outside of the specified band). For the application at hand, these limits have been taken as 200 Hz and 4.8 KHz, with a sampling rate of 10 KHz. The number of delay line stages has been selected as 2N = 30, which results in a maximum amplitude error between the frequency limits of less than 2% using weights given in Reference 17.

5.3 Change Coupled Device (CCD) Implementation

The configuration shown as Figure 13 may be readily implemented using a CCD shift register. A distinct advantage of this approach is that analog-to-digital and digital-to-analog converters (ADC and DAC) are

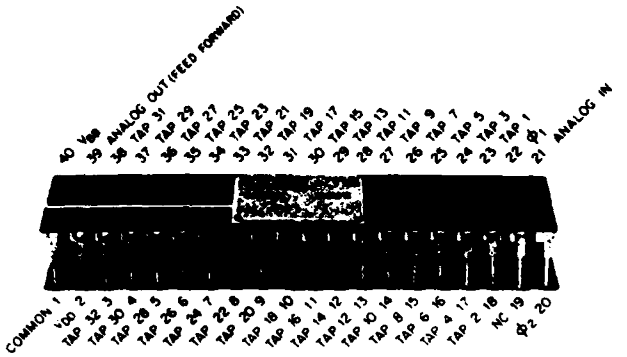
[17] Rabiner, L. R., and R. W. Schafer, "On The Behavior of Minimax FIR Digital Hilbert Transforms," BSTJ, Vol. 53, No. 2, February 1974.

not required as the speech samples are stored as charge packets proportional to sample amplitude. Further, the low power consumption and small size of the CCD technology makes the use of CCDs attractive for portable radio applications. A CCD shift register is commonly referred to as an sampled data analog delay line.

With a shift register length of 2N = 30, a Reticon TAD-32 CCD delay line has been used to realize the HT. This delay line has 32 stages or taps, with the first weighted output being taken from tap 1, and the last weighted output is taken from tap 31. The output corresponding to the delayed version of the original input signal is taken from tap 16. Figure 14 shows the key features and specifications of the TAD-32.

Reticon supplies a TC-32A evaluation card which provides 2-phase clock driver circuits, bias voltages, and a tap weight network operated in conjunction with a differential summing amplifier. The initial evaluation of the CCD HT was made using a TC-32A-02 card which allows the tap weights to be selected by means of 1KΩ, 15-turn, trimpots. Although the tap resistance of each trimpot may be easily calculated for each desired weight (with 0.5KΩ representing W = 0), setting each trimpot to the desired value is difficult because of their parallel in-circuit connections. A network analysis in conjunction with computer based calculations allows each weight to be set by connecting a digital ohm-meter to each tap and adjusting the trimpot to the specified value. Accuracy is good for large weighting values (≥ 0.2), but is poor for the smaller weights. As a result, after experimentation with the TC-32A-02 showed that the CCD delay line was operating properly in terms of circuit parameters, a TC-32A-03 circuit card was then used which allows the use of fixed resistors to be added to very accurately set the tap weights. With 5% resistors used to set the tap weights, the overall passband accuracy has been measured to be within ± 0.5 dB (± 11%) of the ideal all-pass characteristic.

RETICON® TAD-32 TAPPED ANALOG DELAY



The Reticon TAD-32 is a tapped analog delay line fabricated with the most advanced n-channel silicon-gate integrated-circuit technology. It consists of a charge-transfer device with 32 taps equally spaced one sample-time apart along the device. It is designed specifically for use in the realization of transversal filters, but it likewise is applicable to recursive or other filter types. Typical applications include: low pass filters, band pass filters, matched filters, phase equalizers, phase shifters, tone generators, function generators, correlators, and simple tapped delays.

KEY FEATURES

- **Monolithic construction**
- **Full wave or boxcar output from each tap**
- **32 equally spaced taps, with separate feed-forward tap**
- **Buffered outputs from each tap**
- **Tap delay linearly variable with clock period**
- **Sampling rates to 5 MHz**
- **40 db passband-to-stopband ratio (as a filter)**
- **60 db dynamic range**
- **Simple I/O and clock circuit**
- **Low power dissipation**
- **40-pin dual-in-line package**

SPECIFICATIONS (25°C)

Absolute Maximum Rating

	Min.	Max.	Units
Voltage on any terminal with respect to common	-0.4	+20	Volts
Storage temperature	-55	+125	°C
Temperature under bias	-55	+85	°C

Drive

	Min.	Typical		Units
Clock frequency	0.001		5	MHz
Clock amplitude, $V_{\emptyset}$ (Figs. 1, 2)	10	15	16	Volts
Clock line capacitance (each)		50		pf
V_{bb} (optimum)		$V_{\emptyset}$-1	$V_{\emptyset}$	Volts
V_{dd}	$V_{\emptyset}$	+15	16	Volts
DC power dissipation*		200	700	mwatts

*DC power dissipation is strongly dependent on the number of taps used and on tap load currents. When all 32 taps are used with 10 K ohm loads, typical dissipation is 200 mwatts

Input/Output

	Typical	Max.	Units
Number of taps	32		
Input capacitance @ +4 V Bias	8		pf
Output capacitance of each tap @ +5 V Bias	3		pf
Output transconductance (at +5 V level, 10 K Ω d-c load)	1.1		ma/v
Input bias	+4.5		Volts
Input signal (p-p)		4	Volts
Tap d-c level	+5		Volts
Unused taps	Connect to V_{dd}		

Performance Characteristics

	Typical	Max.	Units
A. Single-tap response:			
Dynamic range (See Figs. 3 and 8)	60		db
Linearity (See Fig. 8)			
Harmonic intercepts (See Fig. 8)			
B. 32-tap summed response:			
Dynamic range (See Fig. 3)	60		db
Input sensitivity, S/N =1	4		mvolts p-p

FIGURE 14 - RETICON CCD SPECIFICATIONS

Figure 13 shows input and output lowpass filters (LPF) used respectively for anti-aliasing and interpolation. The filter type selected to perform these functions is a National Semiconductor AF132 Dual PCM Transmit/Receive Filter. Both filters have third order eliptic characteristics. The input or sampling LPF has flat response from 300 Hz to 3 KHz, a -3 dB frequency of 3.8 KHz, and a -50 dB notch at 5.8 KHz. The output or interpolation LPF has a response which compensates for the inherent sin(f)/f sampling function degradation associated with rectangular pulse sample representations. (Note, prior to the use of this filter type, the outputs from the CCD HT had to be reclocked using very narrow sampling pulses so as to eliminate clock switching spikes and minimize interpolation distortion. The AF132 filter eliminates the need for reclocking.) Figure 15 shows the measured (using white noise in conjunction with the HP-3582A spectrum analyzer) response of both filter sections.

Experimental evaluation of the CCD HT has indicated that best performance in terms of minimum distortion and maximum signal to circuit-and-clock noise ratio is obtained when the input speech waveform is scaled to 2 volts peak-to-peak. To maintain this condition over the speaking population dynamic range, a National Semiconductor AF104 AGC/ALC is employed. The overall performance of the CCD HT has been judged as excellent. Unlike the wideband 90° phase-shifters discussed in subsection 5.1, the integrity of the delayed version of the input signal is maintained. The CCD HT has also been incorporated into the Envelope Normalization Breadboard.

The Reticon TAD-32 tapped CCD delay line, operating in conjunction with the TC-32A evaluation card, has provided an expedient means for realizing a CCD based HT in this program. For follow-on developments, and especially for EN implementations used with operational radio systems, a more proficient design is envisioned. A method of incorporating the tap weights directly into the CCD Structure (thereby

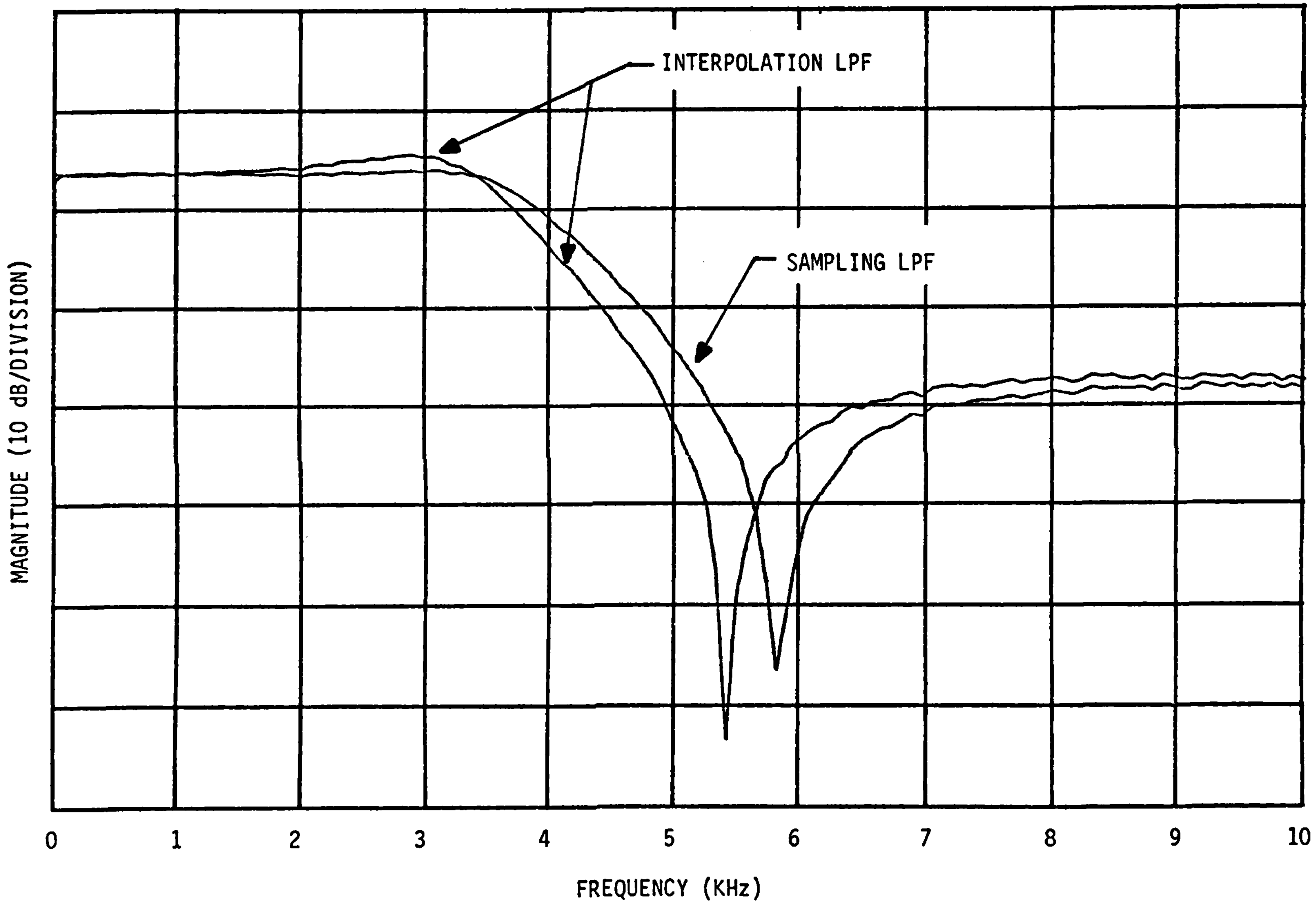


FIGURE 15 - AF132 MAGNITUDE RESPONSES

obviating the use of tap weight resistors) is known as the split-electrode technique.[18] This method is being employed for production line quantities of transversal type filters (which is a prototype of the FIR HT) by several manufacturers. The basic cost involves preparation of the necessary photomasks for the electrode deposition process, following which production may be obtained at low cost (tens of dollars per unit). Thus, the entire configuration of Figure 13 (exclusive of the LPFs) can be fabricated in a small DIP package. For further details on the split-electrode method, consult Reference 18, pp 124-129, and the Reticon data bulletin on thier R5602 Transversal Filter Family.

5.4 Digital Implementations

The analog wideband 90° phase-shifter and sampled data FIR/CCD Hilbert transform mechanizations are highly attractive for the purpose of speech envelope derivation because of their relative simplicity. Digital implementations, on the other hand, are reasonably complex because of the need for an ADC and DAC, and a digital processor in the form of a computer or dedicated logic (or hybrid) to perform the necessary computations leading to the speech HT. An advantage of digital methods is their capability for very high accuracy (not specifically required for speech envelope derivation). Probably the most attractive attribute of the digital approach is that all computations needed to produce an EN signal prior to DAC and interpolation may be carried out in the digital realm. With the analog HT's, the envelope computation and speech normalization functions must be implemented using additional analog circuits.

Digital HT's may be obtained by emulating the wideband 90° phase-shifter or the FIR direct approach. The former is realized using the

[18] Howes, M. J., and D. V. Morgan, Charge-Coupled Devices and Systems, John Wiley & Sons, 1979.

basic computational configuration shown in Figure 16, which has the sampled data transfer function

$$H_i(z) = \frac{z^{-1} - p_i}{1 - p_i\, z^{-1}} \tag{56}$$

corresponding to the difference equation

$$y_n = x_{n-1} + p_i\, (y_{n-1} - x_n). \tag{57}$$

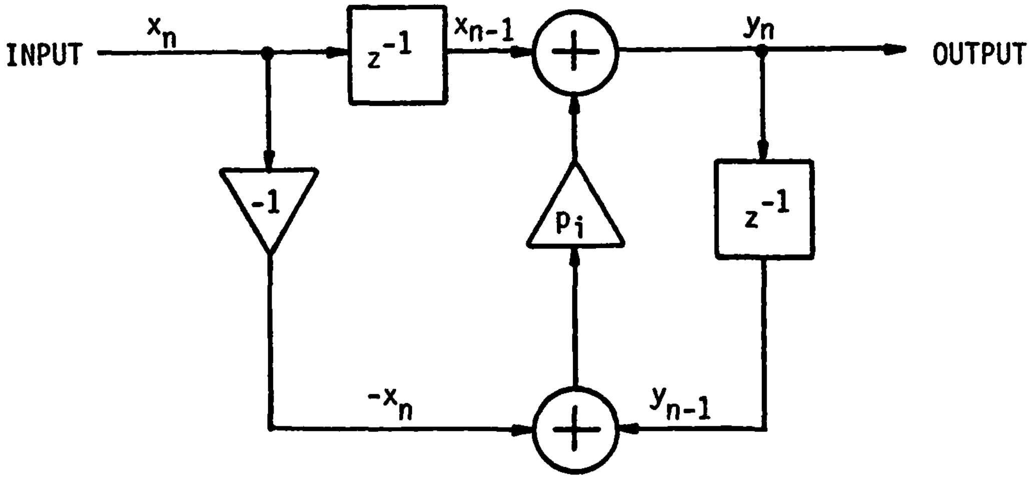


FIGURE 16 - SAMPLED-DATA FIRST-ORDER ALL-PASS SECTION

Each digital algorithm of this form is equivalent to the active all-pass section of Figure 12a, with p_i being specified in a manner akin to α_i. Thus, only six sets of computations are required for the digital emulation of Figure 12b, requiring a total of 6 multiplications and 12 additions. (These are the most time consuming operations.) For the immediate application, these operations might be accomplished using a very fast microprocessor (e.g., the TI TMS99000), with a sampling rate on the order of 6-8 KHz.

The FIR HT in the digital realm is a digital computational equivalent of Figure 13. Because of the odd symmetry of the weights, the number of required multiplications is 8, and the additions total 16, in order to obtain an HT with 2N = 30. This results in too many operations for even a very fast microprocessor to accomplish in real-time, thus, dedicated arithmetic logic must be used. This approach is generally large in size, consumes a moderate amount of power, and is costly.

Recently, Western Digital Corporation announced the availability of a programmable digital filter in the form of a 40 pin DIP which performs all of the transversal operations of Figure 13. Using NMOS technology, this unit (Model WD3150) has the following capabilities and features.

Signal sampling rate:	up to 12 KHz
Number of coefficients (taps):	256
Input bits:	10 plus sign
Output bits:	11 plus sign
Coefficients:	totally programmable
Inputs and Outputs:	TTL compatible
Power Supply:	+ 12 V @ 55 ma + 5 V @ 110 ma

Internally, the unit uses an A-law companded multiplication algorithm. It is believed that if a digital implementation of EN is desired, then the WD3150 operating in conjunction with a microprocessor offers the best solution. Data on the WD3150 was obtained very late in this program, and for this reason, plus the belief that a digital implementation of EN becomes excessive relative to the requirements, detailed digital designs were not pursued.

6.0 CIRCUITS FOR SPEECH ENVELOPE NORMALIZATION

6.1 Envelope Calculation and Speech Normalization

Figure 17 shows the functional EN circuits. Hilbert transforms are discussed in Section 5.0, and the present subsection is concerned with the envelope calculation and the division of the speech by its envelope. Several designs were evaluated during the program.

The original breadboard was implemented using Analog Devices AD533 multipliers as the core elements for squaring, square-rooting, and dividing. In order to obtain optimum performance, each squaring function requires three trimpots, while the square-root and division functions require four trimpots each. Setting of these trimpots for best overall performance was found to be somewhat difficult.

The squaring functions are most easily adjusted; setting the output for zero direct voltage offset (with the input zero), and minimizing input signal feedthrough. The gains are set to give the same output level for a common input. Once set, these circuits appear quite stable and drift free, and therefore cause no further problems.

The square-root function is the most critical to calibrate and maintain. The desired transfer function is

$$v_r = -\sqrt{10v_s} \quad , \tag{58}$$

with v_s ranging from 10mV to 10V. Adjustments set the output relationship for v_s values of 0.1V, 2V, and 10V. The biggest problem is with the low values of v_s. Exact calibration for v_s = 0.1V usually results in an output offset of about 0.2V when v_s = 0. This is undesirable as it limits the divisor value to the divider circuit. As a compromise, the square-root function output for v_s = 0 is adjusted to less than 20mV. Although this results in some inaccuracy of the transfer function for

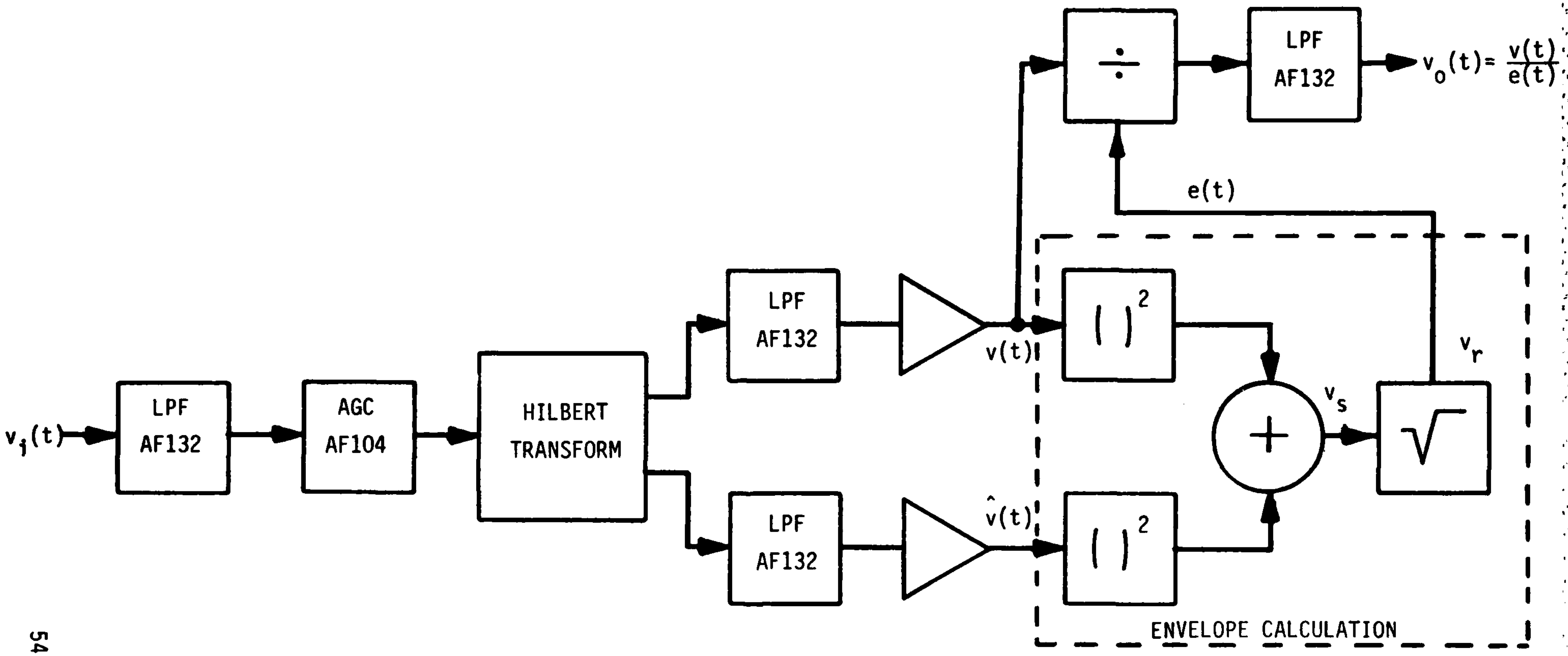


FIGURE 17 - FUNCTIONAL ENVELOPE NORMALIZATION CIRCUITS

small input values, it has been found to have an insignificant effect on the ultimate EN process.

The division circuit is adjusted to minimize output offset and input feed-through. Once properly set, the divider performs well over a 50:1 divisor input range and is reasonably drift free.

Initially, each of the above functions were independently aligned using a fixed sinewave signal together with direct voltage inputs. When they were then operated together to form the entire EN function, with a sinewave input to the Hilbert transform, apparently good EN performance was observed on the scope as an essentially constant sinusoidal output voltage when the sinewave input amplitude was manually increased/decreased.

Following the initial EN circuit alignment, a speech signal was input to the breadboard. The observed results (on the scope) showed imperfect EN operation and some output peak level clipping. Investigations showed that the envelope waveform from the square-root circuit was compressing near the 20mV level. Also, the squaring circuits were biased slightly to the negative side of zero volts.

Following some further experimentation, it was decided that a good EN signal could be obtained by performing a dynamic alignment (i.e., with a speech signal present) of the EN functions. Relationships (24) and (25) form a basis for the dynamic alignment of the squaring and square-root circuits. Further analysis showed that when the gains and biases are improper, (24) and (25) will be violated, and outputs less than zero will be obtained. Thus, by observing the sums $v(t) + e(t)$ and $\hat{v}(t) + e(t)$, dynamic adjustment of the squaring and square-root circuits are made to eliminate any negative output, resulting in a proper $e(t)$ waveform.

With a good $e(t)$ signal, the EN output was again observed and found to be significantly improved. Dynamic adjustment of the divider, especially the divisor input offset and gain, were instrumental in producing excellent results.

Although it was possible through the just outlined procedures to obtain a satisfactory EN waveform using AD533's, the large number of critical adjustments required is highly undesirable, and certainly must be eliminated for any operational design. As a result, the breadboard design was modified to incorporate two new devices. The first is a National LH0094 multifunction converter, which, when connected as shown in Figure 18, performs the operation of calculating vector magnitude, i.e.,

$$v_r = \sqrt{v_1^2 + v_2^2}\ , \tag{59}$$

without the need for any trim adjustments (apart from equalizing the RMS amplitudes of v_1 and v_2). For small values of v_1 and v_2, the expected error is less than 0.5 % of full scale (i.e., less than 50 mV). The absolute value circuits each make use of an LF353 dual FET amplifier, two diodes, and five identical resistors in the form of a resistor DIP. An LF353 and resistor DIP are also used for the noninventing summing amplifiers. As a result, the entire vector magnitude converter is constructed with an absolute minimum of discrete parts, and has no trimming adjustments.

The arithmetic performance of the vector magnitude converter is within ±0.5% of the true value over a 48 dB input range (20 mV to 5V). Actually, the accuracy is limited by the ability to balance the amplitude of the two inputs as well as being able to maintain them in perfect phase quadrature. This accuracy is more than acceptable for the EN function.

The only departure from near ideal performance on the part of the multifunction converter is that the output contains some harmonic components of the input signals. The problem is that the transfer characteristics from the three input ports to the output of the LH0094 have a frequency response which is a function of the input voltage level. Notice from Figure 18 that the input to the divide port is $|v_2| + v_r$. Consider the

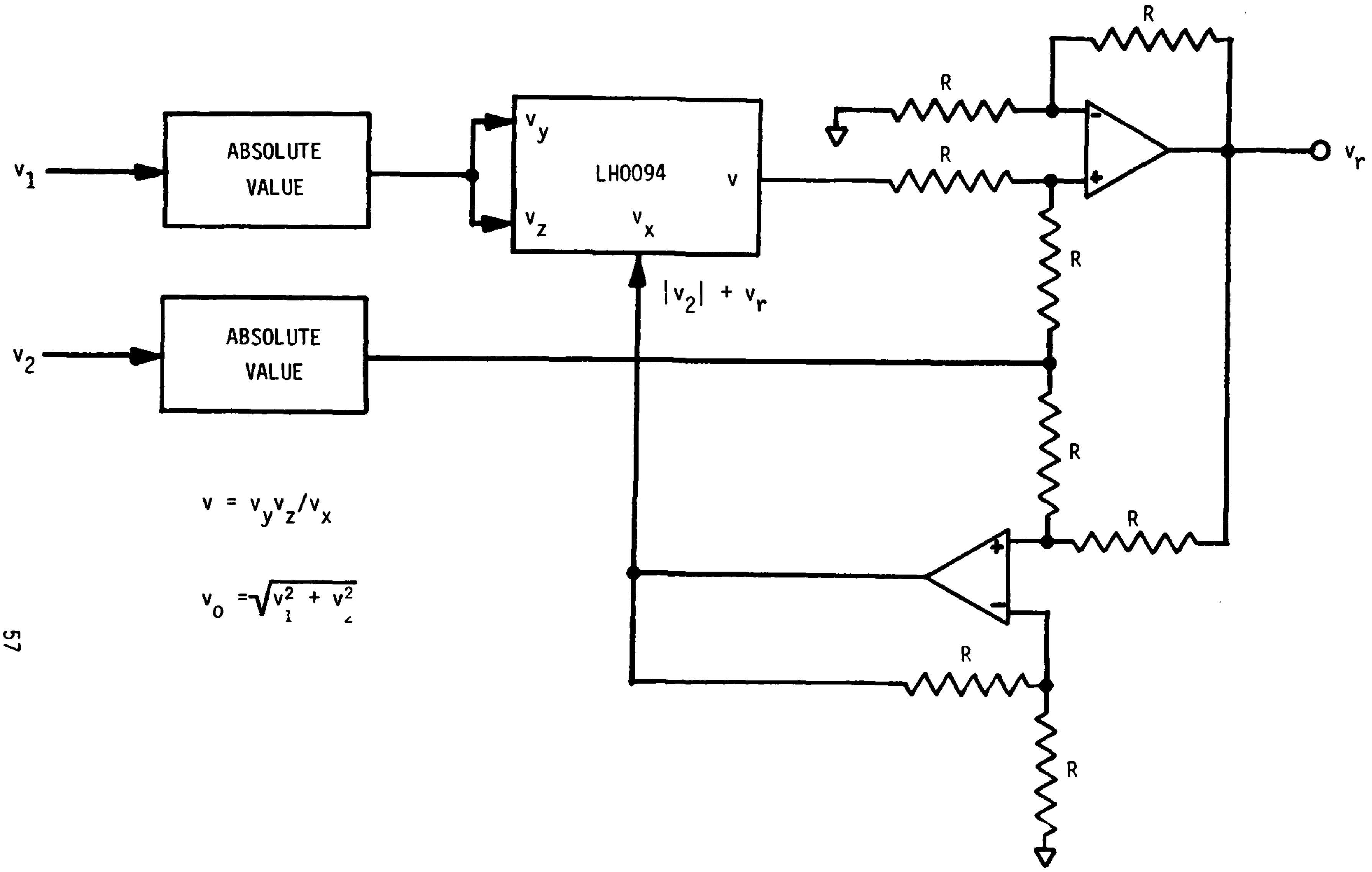


FIGURE 18 - VECTOR MAGNITUDE CONVERTER

two inputs to be quadrature sinewaves. Then ideally v_r is a constant while $|v_2|$ is rich in even harmonics of the sinewave frequency. If the LH0094 attenuates and phase-shifts these harmonics by acting as a low-pass filter, then the output v_r must be a constant plus even harmonics of the sinewave frequency. Typically the LH0094 has a -3 dB frequency on the order of 10-15 KHz. As a result, the harmonic content of the output when f = 3 KHz can be appreciable. Figure 19 is a graph of the ratio of the multifunction converter output direct voltage level divided by the RMS value of the harmonic produced waveform (ratio expressed in dB) vs. frequency. This graph was prepared from measurements made on the multi-function converter breadboard. As may be seen, the harmonic content of the envelope function increases directly with frequency. A 20 dB ratio represents a 10% total harmonic content.

At first the presence of the signal harmonics was of concern. However, experiments quickly established that when the output of the envelope normalization divider is lowpass filtered, the effect of the harmonics, i.e., EN waveform distortion, is essentially eliminated. This is because in the upper frequency range (>1.5 KHz) where the harmonics become significant, the distortion products fall largely outside of the lowpass filter. On the lower frequency range (<1.5 KHz) where the distortion products fall inband, their amplitude is sufficiently small that they do not cause a problem. Even without the lowpass filter, for speech signals the subjective effect of the harmonic distortion products is to appear as a high frequency rasp, which although noticeable (when a filter-in filter-out comparison is made), is not particularly objectionable, especially compared to the raspiness introduced by the basic EN process itself (see subsection 9.3).

Even though the vector magnitude converter performs quite satisfactorily in spite of the harmonic content problem, a third approach for envelope production was breadboarded and evaluated. Figure 20 shows the functional approach. The two baseband signals, v(t) and $\hat{v}(t)$ are balanced-modulated onto quadrature carriers of frequency f_c.

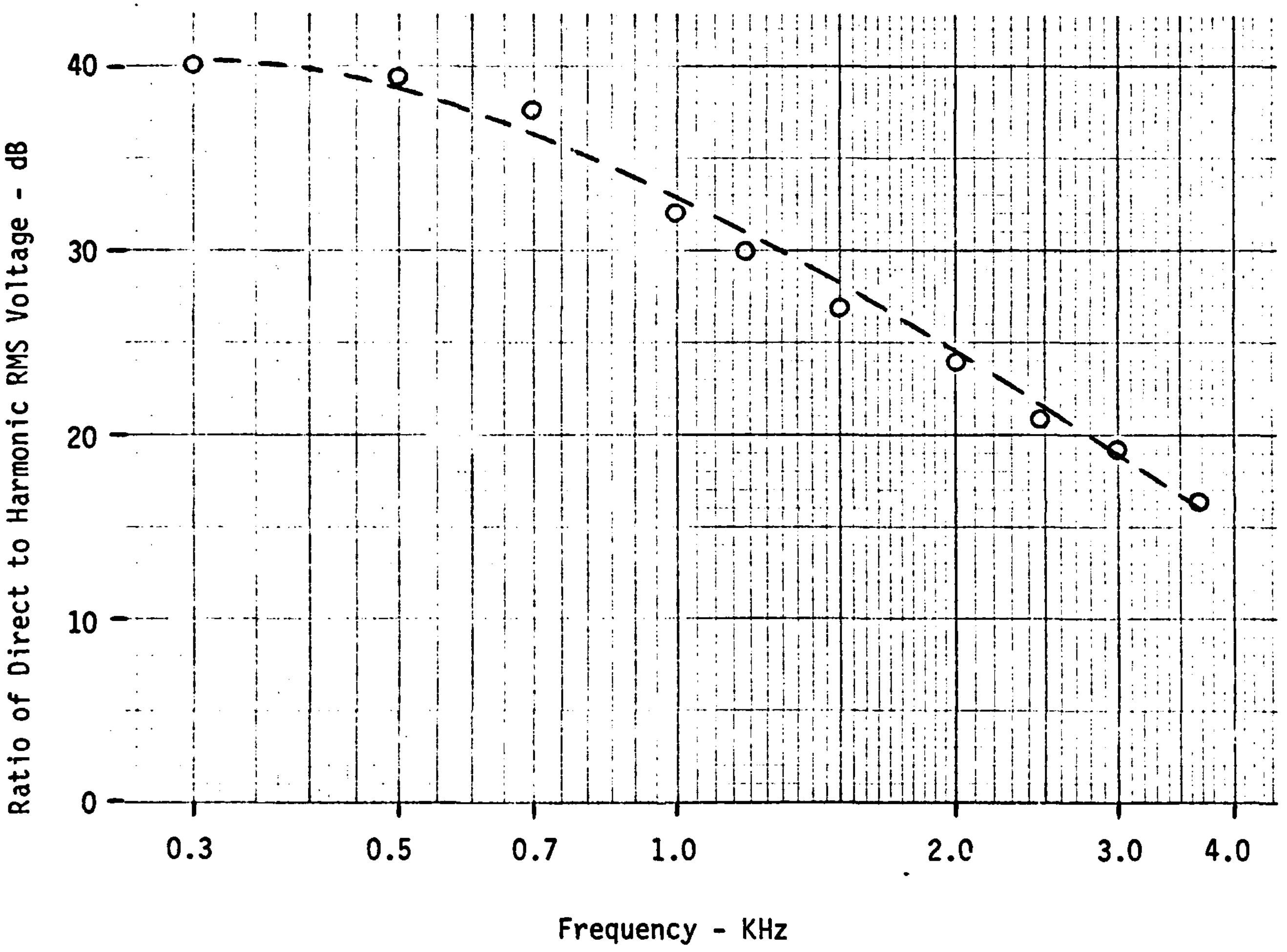


FIGURE 19 - LH0094 VECTOR MAGNITUDE CONVERTER HARMONIC OUTPUT

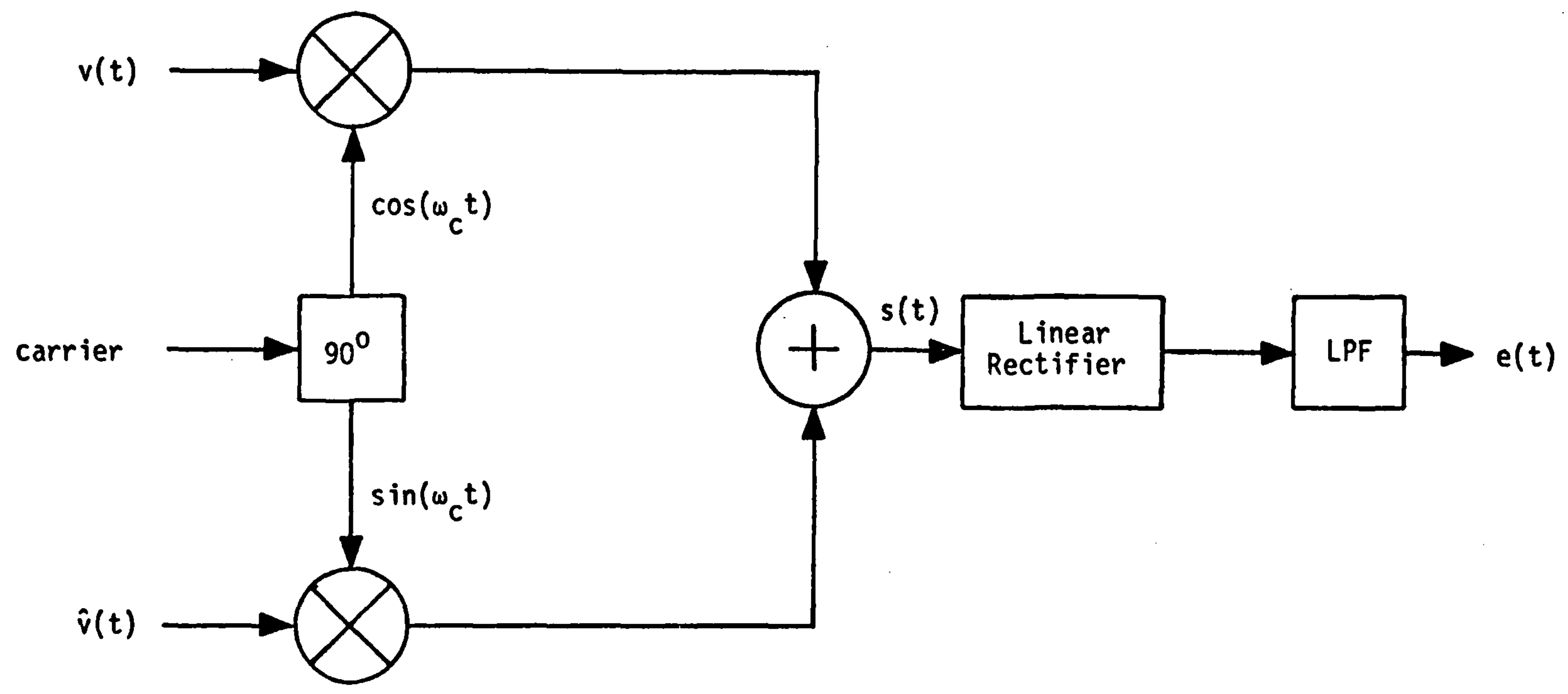


FIGURE 20 - ENVELOPE DERIVATION USING A MODULATOR/LINEAR-RECTIFIER

The results are added to form the signal

$$s(t) = v(t)\cos\omega_c t + \hat{v}(t)\sin\omega_c t = \sqrt{v^2(t)+\hat{v}^2(t)}\cos\left[\omega_c t + \psi(t)\right]. \quad (60)$$

This signal is then input to a linear rectifier and lowpass filtered to eliminate the carrier harmonic, $2f_c$. The result is the desired envelope.

For the breadboard mechanization, LM1496 balanced modulators were used. The quadrature carriers are squarewaves (rather than sinewaves) derived from 180°-phase-clocked flip-flops. A carrier frequency (squarewave fundamental frequency) of 250 KHz was employed. The linear rectifier is the same basic absolute value circuit used with the vector magnitude converter.

Measured performance is very good. There is essentially no harmonic content from the input signals in the resultant envelope, and the dynamic range is on the order of 40 dB. The principal problem is that the modulators produce a form of switching noise which appears added with the desired envelope waveform. As a result, the envelope to circuit-noise ratio is poor when the envelope level is small. It is believed, however, that this problem can be minimized when the circuit is properly constructed on a ground-plane circuit card (rather than a Proto-board).

A second problem is that because the output of the linear rectifier must be filtered, a delay of the envelope waveform ensues, and this same delay must also be introduced into the speech signal to the divider. Measurements have shown this compensation to be quite critical if an acceptable EN waveform is to be obtained. Lastly, the approach requires a moderate amount of power to properly operate the LM1496 modulators.

After careful consideration of the advantages and disadvantages of the different methods of deriving e(t) from v(t) and $\hat{v}(t)$, it was decided to incorporate the vector magnitude converter into the demonstration breadboard.

The second device incorporated into the modified breadboard was an Analog Devices Model 436 precision divider in place of the AD533 divider. This divider, which uses log/antilog circuits in conjunction with a variable transconductance differential amplifier, has an accuracy of better than ± 0.5% over a divisor range of more than 100:1 without the need for any external trimming. With optional external trimming of the divisor offset, the accuracy can be extended to a divisor range of 1000:1. Although trimming is generally not necessary, provision to slightly offset the divisor input has been made to keep the divider output from saturating at ±10v when the envelope signal is zero. The dynamic range of the divider under this condition is 54 dB.

In summary, the LH0094 vector magnitude converter and Model 436 divider perform the EN operations in the EN demonstrator. Measured performance is presented in subsection 9.1.

6.2 VOX Configurations

The need for a VOX to switch-out large noise bursts caused by the EN process during speech pauses has been outlined in subsection 3.3. VOX mechanization will now be considered. In order to perform this operation reliably, so that weak speech segments will not be lost, a speech vs. silence discriminator based upon an algorithm suggested by Rabiner and Schafer[19] was designed. Figure 21 shows the functional configuration.

The speech vs. silence discriminator design contains two detectors, an energy detector and a zero crossing detector. The zero crossing detector is used primarily to detect the presence of weak unvoiced segments. Operation of both detectors is based on a finite observation period, T_s, and if either detector has an output at the end of T_s which exceeds its respective preset threshold, then voice presence is declared. Initially, T_s = 10 ms.

[19] L. R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, 1978.

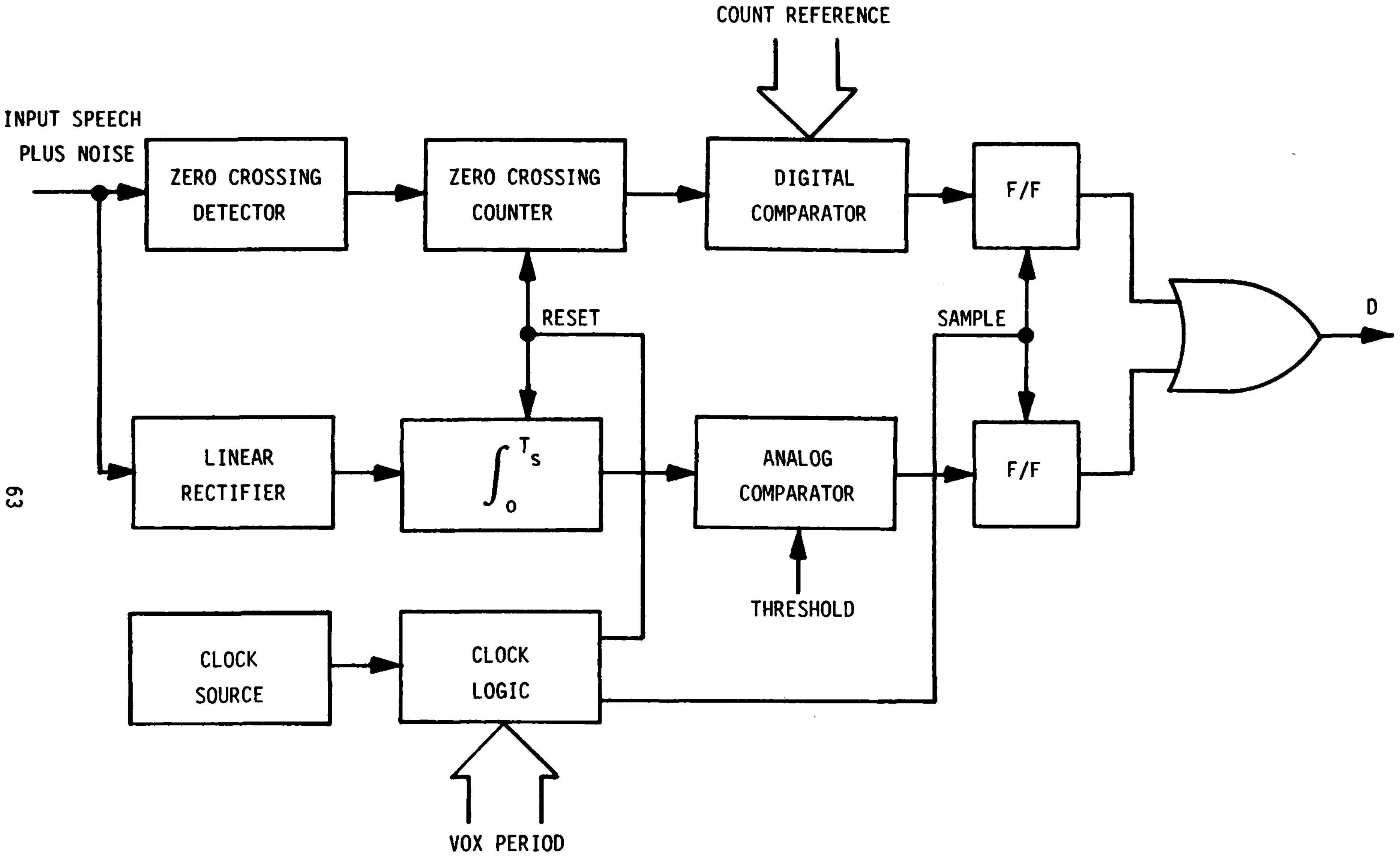


FIGURE 21 - SPEECH VS SILENCE DISCRIMINATOR FUNCTIONAL DIAGRAM

The energy detector is a linear full wave rectifier followed by an integrator which is reset or discharged to zero at the end of T_s. Just prior to being reset the integrator output is compared with a reference voltage using an LM211 comparator. Since the reference is fixed (once the correct value has been established), the input to the energy detector is taken from the AF104 AGC regulator in order to provide acceptable performance over a large speech dynamic range. The initial design of the energy detector when tested was found to perform as expected.

The zero crossing detector consists of an LM211 comparator acting as a limiter, followed by a pair of SN74121 monostable multivibrators which produce pulses respectively for positive and negative zero crossings. These pulses are counted by two SN74193 binary synchronous counters, with the outputs being compared with a fixed number using two SN7485 magnitude comparators. When the count exceeds the fixed number, speech presence is declared. Again, these circuits initially performed as designed.

Experiments established that the ability of the zero crossing detector to discriminate against wideband noise was not as good as desired. The expected number of positive plus negative zero crossings in a 10 ms period for unvoiced speech is on the order of 50.[19] On the other hand, for Gaussian noise which is essentially flat on the range of 300 Hz to 3 KHz, the expected number of positive plus negative zero crossing in a 10 ms period is 37.[20] Because the standard deviation on the speech zero crossings is about 10, while that for the noise is on the order of 5, there is significant overlap in the two distributions of zero crossings. As a result, there is not as sharp of a demarcation between unvoiced speech and noise as was hoped for initially. Based upon experiments, it has been found that the zero crossing detector responds

[20] Rice, "Mathematical Analysis of Random Noise," BSTJ, Vol. 24, 1945.

to noise more often than desired and is therefore a detriment to the overall VOX function. The zero crossing detector, therefore, was deleted from the VOX design.

Basically, the energy detector is intended to detect the stronger speech segments which are considerably greater than the noise floor. However, with only the energy detector remaining in the speech vs. silence discriminator, experiments were conducted to determine the best value of T_s with respect to the fixed threshold reference for minimization of lost (undetected) speech segments. A value of T_s between 2 ms and 5 ms appears to give the most satisfactory results. With T_s = 2 ms, it is not necessary to incorporate a specific speech delay circuit, as the combination of filters and CCD Hilbert Transform (or wideband 90° phase shifter) between the input and the EN divider jointly provide about this amount of delay. Thus, in the interest of circuit economy, T_s has been set at 2 ms.

Performance of the VOX in conjunction with the EN signal appears in subsections 9.1 and 9.3.

7.0 OTHER POTENTIAL ENVELOPE NORMALIZATION APPROACHES

At the onset of the program, it was hoped that some clever method of producing EN signals without the need for directly forming the Hilbert transform, envelope, and division operations might be devised. Several ideas were pursued, but after some initial study, it was concluded that any technique that does not involve speech signal delay processing equivalent to the HT, will not generate satisfactory results. As pointed out in subsection 3.2, even the RF clipping method involves such delay by virtue of the need for a multipole sideband BPF. Two alternate methods were, however, given some detailed consideration, and are summarized in the following subsections.

7.1 Sine-Pulse Approximation

The idea for a sine-pulse approximation to the speech EN waveform is suggested by eqn. (22), and supported by temporal records of EN speech (see subsection 9.1). The approach is to measure the successive zero-crossing intervals of the speech waveform, and to construct a sine-pulse for each interval, with the polarity of each pulse atternating in sign. The concept appears simple, but implementation is found to be complex. Figure 22 shows a functional digital configuration. The principal problem stems from the fact that each zero-crossing interval of the speech waveform must be measured and stored for a variable period of time. That this is so may be seen from the waveform shown on Figure 22 where the period T_1 is much larger than the few periods which follow. Since the algorithm requires T_1 seconds to synthesize a sine-pulse of duration T_1, the periods T_2, T_3, and T_4 must be measured and held until they are needed. It can be seen, in general, that a reasonable number of periods must be available in the memory so that the memory never becomes empty, and the memory must be large enough so that overfill does not occur. Memory input and output is asynchronous, therefore a first-in, first-out, (FIFO) type of memory is required. A microprocessor operating in conjunction with a frequency synthesizer (actually a down-counted high frequency clock) and a ROM of sine-pulse values, synthesizes the sine-pulse of proper duration.

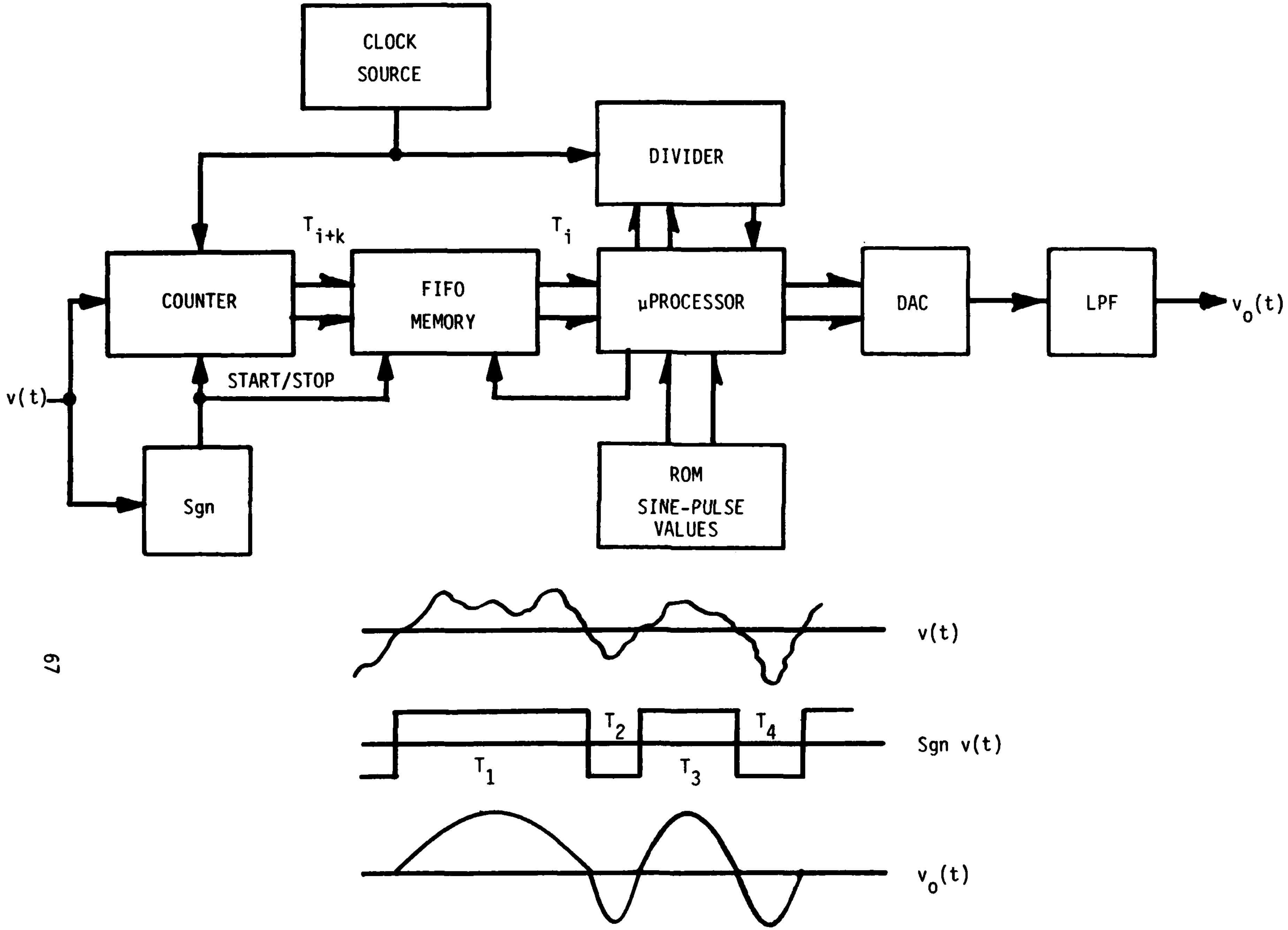


FIGURE 22 - SINE-PULSE APPROXIMATION OF EN SPEECH

An analog circuit equivalent of the Figure 22 functional configuration appears unfeasible. The digital mechanization is not simple, and certainly cannot rival the digital realizations of the direct approach for producing an EN signal.

One possible advantage of the sine-pulse approximation is that it virtually eliminates additive noise when speech is present. On the other hand, speech distortion could prove excessive, particularly since small inflections in the true EN waveform will also be removed. Since the method has not been implemented, the subjective performance remains unknown.

7.2 Instantaneous Frequency Generator

Another potential method for generating an approximate EN speech signal is to make use of a frequency-to-frequency converter where the instantaneous frequency of the speech becomes the instantaneous frequency of a sinewave. The concept consists of a frequency-to-voltage converter followed by a voltage-to-sinewave converter. A promising voltage-to-sinewave converter (VCO), capable of operating over a frequency range of 200 Hz to 4 KHz, was breadboarded and evaluated with excellent results. This, therefore, encouraged further study into frequency-to-frequency converter implementations.

The specified performance of several commercially available frequency-to-voltage converters was reviewed. All of these converters operate on the principle of zero-crossing-rate to voltage transformation. Unfortunately, none of the converters studied has sufficient frequency response to follow and produce a credible output representative of the instantaneous frequency of the speech signal.

The basic problem of frequency-to-voltage conversion can be understood from the equation which gives the exact instantaneous frequency of any lowpass signal, x (= x(t)), viz.,

$$\text{Instantaneous Frequency} = \frac{x\dot{\hat{x}} - \hat{x}\dot{x}}{x^2 + \hat{x}^2}, \qquad (62)$$

where the over-dot denotes the time derivative. Because (62) involves $\hat{x}$, and any good estimate of $\hat{x}$ involves delay (or memory), it is easy to see why the aforementioned frequency-to-voltage converters do not meet the needed requirement; they do not embrace any delay-based processing (or functions). If an acceptable frequency-to-voltage converter is to be realized, it must involve delay on the order of that needed to produce credible Hilbert transforms. However, since the hoped for objective of the frequency-to-frequency converter is to avoid the need for implementing the HT (and the EN functional processing), it appears that wishes and reality cannot be equated. Therefore, a proper frequency-to-frequency converter, which is simpler than the direct mechanizations for producing EN speech signals, is considered unattainable.

8.0 SPEECH ENVELOPE LINKING

In order to perform expansion at the receiver, the speech envelope used in the EN operation at the transmitter is required. For ν:1 companding (see subsection 2.1 for definitions) the process of expansion may be based solely on the received compressed speech signal (at least theoretically speaking) provided ν is small (i.e., $\nu = 2$ or $\nu = 4$). However, because EN speech has a constant envelope, the EN operation is not self-reversing (just like the process of passing a waveform through a hard limiter is not reversable). Thus, a subsidiary means of transmitting the speech envelope to the receiver must be employed.

Speech envelope transmission is generally effected by modulating the envelope signal onto a subcarrier or pilot tone placed below or above the audio frequency band. A commerical system which employs this approach for 4:1 companding using FM is known as "Lincompex."[21] In another system, the waveform is quantized (typically to 8 levels), and the information is transmitted using FSK modulation of the subcarrier. This commercial system is called "Syncompex." Both of these methods are employed with overseas telephony operating on the HF band. A form of speech envelope transmission using AM on the pilot has been recently demonstrated[9] in conjunction with 4:1 compressed speech. The following subsections examine the theoretical and implementation aspects of envelope waveform transmission with EN speech. The generic operation is referred to as envelope "linking."

8.1 Basic Requirements and Pilot Modulation Methods

The speech envelope is a lowpass signal whose bandwidth is relatively small compared to the bandwidth of speech itself. Little information has appeared in the literature concerning the exact spectral characteristics of the speech envelope, although it has heretofore been reported[22] that the principal frequency components lie on a range of

[21] Turner, L. W., Electronic Engineer's Reference Book, Butterworth & Co., 1976.

[22] Horii, Y., et. al., "A Masking Noise with Speech-Envelope Characteristics for Studying Intelligibility," The Journal of the Acoustical Society of America, Vol. 49, No. 6 (Part 2), 1971.

0 Hz to 25 Hz. Such indeed has been verified by spectrum measurements made in the present program (see subsection 9.2). However, it has also been found that envelope frequencies as high as 300 Hz are important to quality EN speech expansion (see subsections 9.2 and 9.4). Therefore, the linking process should embrace frequencies between 0 Hz and 300 Hz.

Several pilot modulation options have been considered. Generally, analog FM or PM, AM, or digital PCM, of the pilot are the options available. A digital approach does not appear feasible because of the bandwidth required by the serial data stream (sampled and quantized representation of the envelope waveform), and the need for bit and word synchronization for the receiver detection circuits. FM has a distinct advantage in that it intrinsically provides 0 Hz response, while PM (whether coherent or incoherent) does not. AM also has the capability for 0 Hz response, but because the carrier cannot be suppressed (remember, the envelope modulation is unipolar), a direct voltage component at the receiver must be removed. Disadvantages of FM are the need for good pilot and FM detector center frequency stability (in order to minimize received envelope direct voltage offset), and potential modulated pilot bandwidth expansion (which is a function of pilot frequency deviation). AM does not suffer from these detriments. Another very important consideration is that of envelope amplitude control. For AM, some type of AGC will be needed, which complicates the pilot demodulation circuits. Using FM, the needed control is simply supplied by an amplitude limiter. Following additional study of both FM and AM in terms of SNR performance and mechanization tradeoffs, FM was selected for the demonstration breadboard.

8.2 Envelope SNR Requirements

A review of the available literature on speech envelope linking has not disclosed any specific requirement on envelope signal SNR needed for acceptable expansion. As a result, it was decided to specify the envelope SNR into the expandor as a function of expanded speech SNR degradation. Such degradation must be traded-off with regard to the pilot amplitude relative to the EN speech signal (in effect, the transmitter power needed for the linking with respect to that required for the EN speech itself).

Figure 23 is a block diagram of the linking structure, which is appropriate to any form of pilot modulation. The signals e(t) and $v_o(t)$ are given respectively by eqns. (14) and (15), and the corresponding variances by (43) and (44). For an FM pilot, the modulated pilot signal is given by

$$Ks_e(t) = K\cos\left[\omega_p t + k_p \int e(t)dt\right] , \tag{63}$$

where $\omega_p = 2\pi f_p$ is the nominal pilot frequency, and k_p is the pilot FM sensitivity. It is the usual practice to specify the peak (or maximum) frcquency deviation of the pilot, which is given by

$$\Delta f_p = k_p e_{max}, \tag{64}$$

with e_{max} being the maximum value of e(t). If δ is defined as the peak to RMS voltage factor of the envelope, then

$$\Delta f_p = k_p \delta\sigma_e = \sqrt{2} k_p \delta\sigma_v \quad . \tag{65}$$

For speech, $\delta = 3.5$. Notice from (65) that Δf_p is then a function of σ_v, the RMS value of the speech. This relationship will be used subsequently. The task now is to solve for the pilot amplitude, K, required for a stated degradation, X, of the expanded speech SNR.

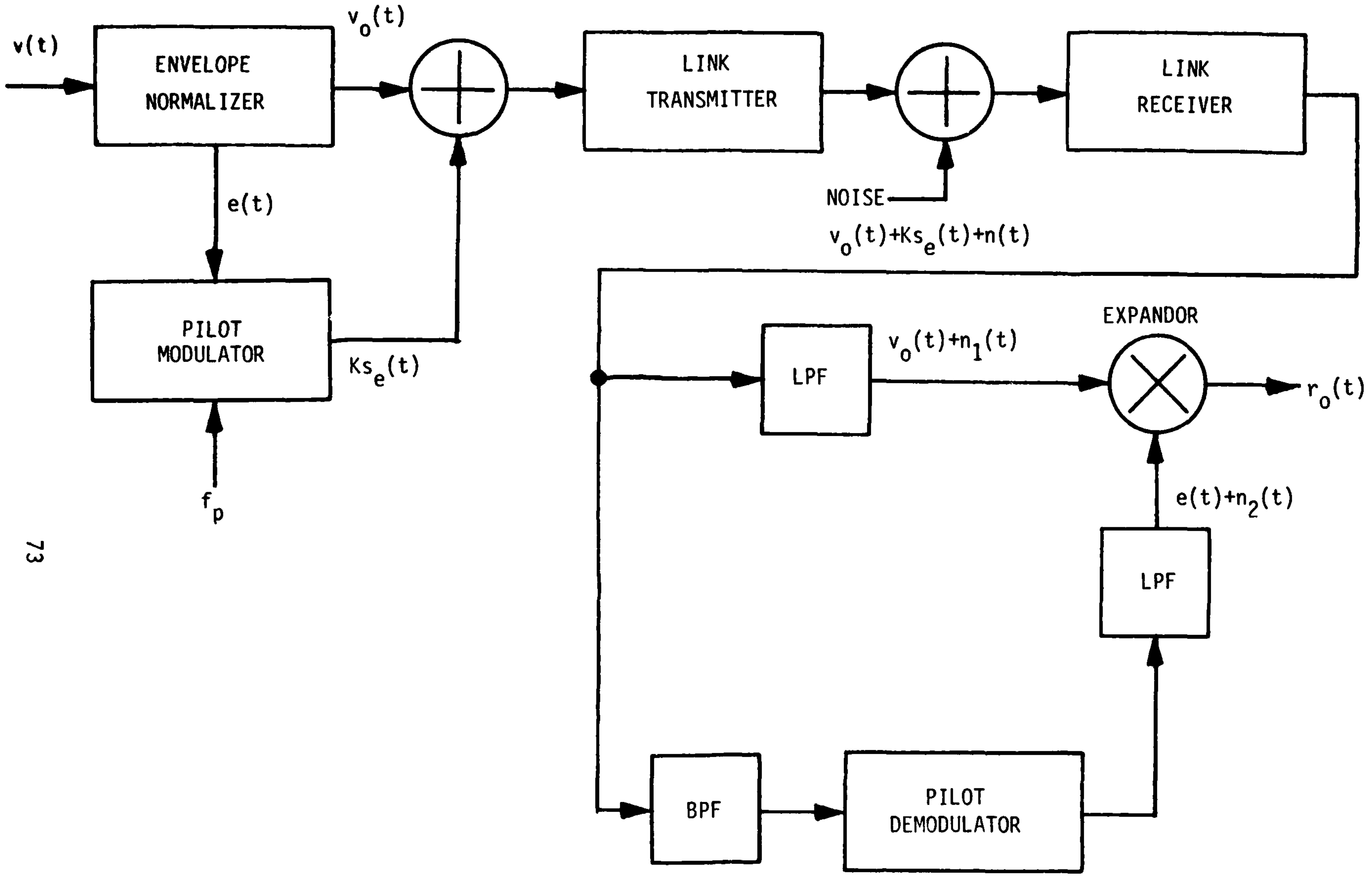


FIGURE 23 - BASIC ENVELOPE LINKING CONFIGURATION

Referring to Figure 23, and using (15), the totality of components at the expandor output is found to be

$$r_0(t) = v(t) + e(t)n_1(t) + v(t)n_2(t)/e(t) + n_1(t)n_2(t). \qquad (66)$$

Of the four components in (66), the first is the desired speech signal, while the remaining three represent noise. The speech SNR into the expandor is defined as

$$SNR_1 = \overline{v_0^2(t)}/\overline{n_1^2(t)} = 1/2\sigma_{n_1}^2 \; . \qquad (67)$$

At the expandor output, the speech SNR is designated by the symbol SNR_2, and using (43) and (44) SNR_2 is found to be

$$SNR_2 = \left[(SNR_1)^{-1} + \sigma_{n_2}^2 \; (1+2\sigma_{n_1}^2)/(2\sigma_v^2)\right]^{-1} \quad . \qquad (68)$$

As may be seen, the result depends only on the two noise variances into the expandor, and the variance of the speech.

It is assumed that the FM pilot demodulator operates above the usual input SNR threshold of 10 dB, thus, standard FM performance relationships[23] may be used. Defining the demodulated envelope SNR as

$$SNR_e = \overline{e^2(t)}/\overline{n_2^2(t)} = 2\sigma_v^2/\sigma_{n_2}^2 \qquad (69)$$

[23] Stremler, F. G., Introduction To Communication Systems, Addison-Wesley, 1977.

this SNR is related to the pilot link FM parameters by

$$SNR_e = (3\rho B\Delta f_p^2)/(\delta^2 f_e^3) \ , \tag{70}$$

where ρ is the SNR in the input bandwidth B (of the BPF) preceding the FM demodulator, Δf_p is the pilot peak frequency deviation, δ is the envelope peaking factor (as previously defined), and f_e is the noise bandwidth of the LPF following the pilot demodulator. Equations (63),(65), (67), (68), (69), and (70) are then combined to solve for the pilot amplitude K. In order to obtain conveniently useful results, two additional relationships are established, viz,

$$\rho = K^2/(2N_oB) \ , \tag{71}$$

and

$$\sigma_{n_1}^2 = N_o f_m \ , \tag{72}$$

where N_o is the effective noise spectral density at the link receiver output (and is a function of the type of link modulation and demodulation employed), and f_m is the noise bandwidth of the EN speech LPF preceding the expandor.

Two results of interest are reported; Case 1 where the link is a baseband additive noise channel (this is representative of the demonstration breadboard operating by itself) and Case 2 where the link employs FM (typical radio link). For Case 1, the pilot amplitude (now designated K_1) is given by

$$K_1^2 = \left[(SNR_1+1)/SNR_1\right](\delta^2 f_e^3)/(3L\Delta f_p^2 f_m) \ , \tag{73}$$

while for Case 2, the pilot amplitude, K_2, is obtained as

$$K_2^2 = K_1^2 \, (3f_p^2)/2\pi f_m^2) \quad . \tag{74}$$

In (73), L is the expandor SNR loss factor, defined as

$$L = X^{-1} - 1 \quad , \tag{75}$$

with

$$X = SNR_2/SNR_1 \tag{76}$$

being the actual speech SNR degradation.

Examination of (73) shows that the result is only weakly dependent upon SNR_1, especially when $SNR_1 > 100$ (20 dB), as is typically the situation, for then

$$(SNR_1 + 1)/SNR_1 \approx 1 \quad . \tag{77}$$

It is also seen that K_2 is a modified function of K_1. In fact, the result for any type of link modulation can be expressed as a modification of K_1, thereby establishing (73) as the foundation envelope linking formula.

In order to assess the results just obtained, values must be assigned to the various parameters. The following list of values are applicable to the demonstration breadboard developed for this program.

$$f_e = 300 \text{ Hz}$$

$$\Delta f_p = 400 \text{ Hz}$$

$$f_m = 3.8 \text{ KHz}$$

$$f_p = 5 \text{ KHz}$$

$$\delta = 3.5$$

Table 3 gives values of K_1 and K_2 as a function of the speech degradation X defined by (76) and expressed in dB for SNR_1 = 20 dB. Thus,

$$SNR_2(\text{dB}) = 20 + X(\text{dB}). \tag{78}$$

X = -1.0 dB	$K_1 = 0.841$	$K_2 = 0.765$
X = -1.5 dB	$K_1 = 0.666$	$K_2 = 0.606$
X = -2.0 dB	$K_1 = 0.560$	$K_2 = 0.509$
X = -2.5 dB	$K_1 = 0.485$	$K_2 = 0.441$
X = -3.0 dB	$K_1 = 0.429$	$K_2 = 0.390$
X = -3.5 dB	$K_1 = 0.385$	$K_2 = 0.350$
X = -4.0 dB	$K_1 = 0.348$	$K_2 = 0.316$
X = -4.5 dB	$K_1 = 0.317$	$K_2 = 0.289$
X = -5.0 dB	$K_1 = 0.291$	$K_2 = 0.265$

TABLE 3 - ENVELOPE LINKING PILOT AMPLITUDES

It is up to the system designer to specify an acceptable value of X or K. Usually, K should be a small fraction of the EN speech signal amplitude, i.e., K << 1. A reasonable assertion is that the pilot power be 10 dB below that of the EN speech, in which case K = 0.316. From Table 3, it can be seen that such results in X = -4.5 dB and X = -4.0 dB degradations for baseband and FM links respectively. It should also be noted that whatever the value of X, the degradation should be applied tc the measurable improvements expected through the use of EN as discussed in Section 4.0.

8.3 Envelope Linking Implementation

Several types of FM pilot modulator and demodulator designs were considered. Principal concern was the minimization of direct voltage offsets due to frequency drift of the modulator and demodulator circuits. An early approach contemplated using an Armstrong type modulator and a quadrature-phase/frequency demodulator in conjunction with crystal controlled frequencies. This however proved unworkable for the pilot peak deviation required (400 Hz). After further study and experimentation, a Wien-bridge oscillator, using a 2N3819 JFET as a voltage variable resistor in order to affect frequency variation, was selected for the pilot modulator. The corresponding demodulator was implemented with a type 565 phase-locked loop. The demodulator input BPF was designed using AF100 state-variable filter circuits.

A nominal pilot frequency of 5200 Hz was chosen to minimize EN speech and pilot signal interference. An interesting feature of using FM is that because the envelope waveform in unipolar, the frequency deviation from the unmodulated pilot frequency is always in one direction, i.e., either above or below the pilot frequency. When EN speech vs. pilot signal interference is considered, using an FM deviation which reduces the unmodulated pilot frequency helps interference discrimination. This happens because, when the speech signal is small, the pilot frequency is furthest removed from the upper limit of the speech band, helping to improve the speech-to-pilot interference ratio (which is a function of the speech and pilot filters). On the other hand, when the speech envelope is close its maximum value and the modulated pilot frequency is near its lowest limit, the decrease in pilot attenuation by the speechband filter is balanced by the greater level of the speech signal.

A final consideration given to the linking problem is that of relative delay between the EN speech and envelope waveforms into the expandor. The main practical difficulty arises due to delays that are inherent in the composite signal separation filters and the process of demodulating and detecting the envelope. As a result, improper time alignment between the EN speech and the envelope is likely. If the transitions between the spoken and silent segments of the EN speech signal do not coincide properly with the dynamics of the envelope, fractions of syllables could be lost, and short intense noise bursts may be perceived by the listener. The situation is illustrated in Figure 24.

To solve this problem, the EN speech to the expandor is delayed in time equal to the differential delay provided by the pilot demodulation filters. The delay line is mechanized using a Reticon SAD-1024 CCD, with the delay being controlled by the clock frequency supplied to the CCD. For the demonstration breadboard, this clock is derived from the master 5.0688 MHz crystal oscillator source, and is adjustable in discrete increments so that delays between 0.8 ms and 6.4 ms, in steps of 0.4 ms, may be selected.

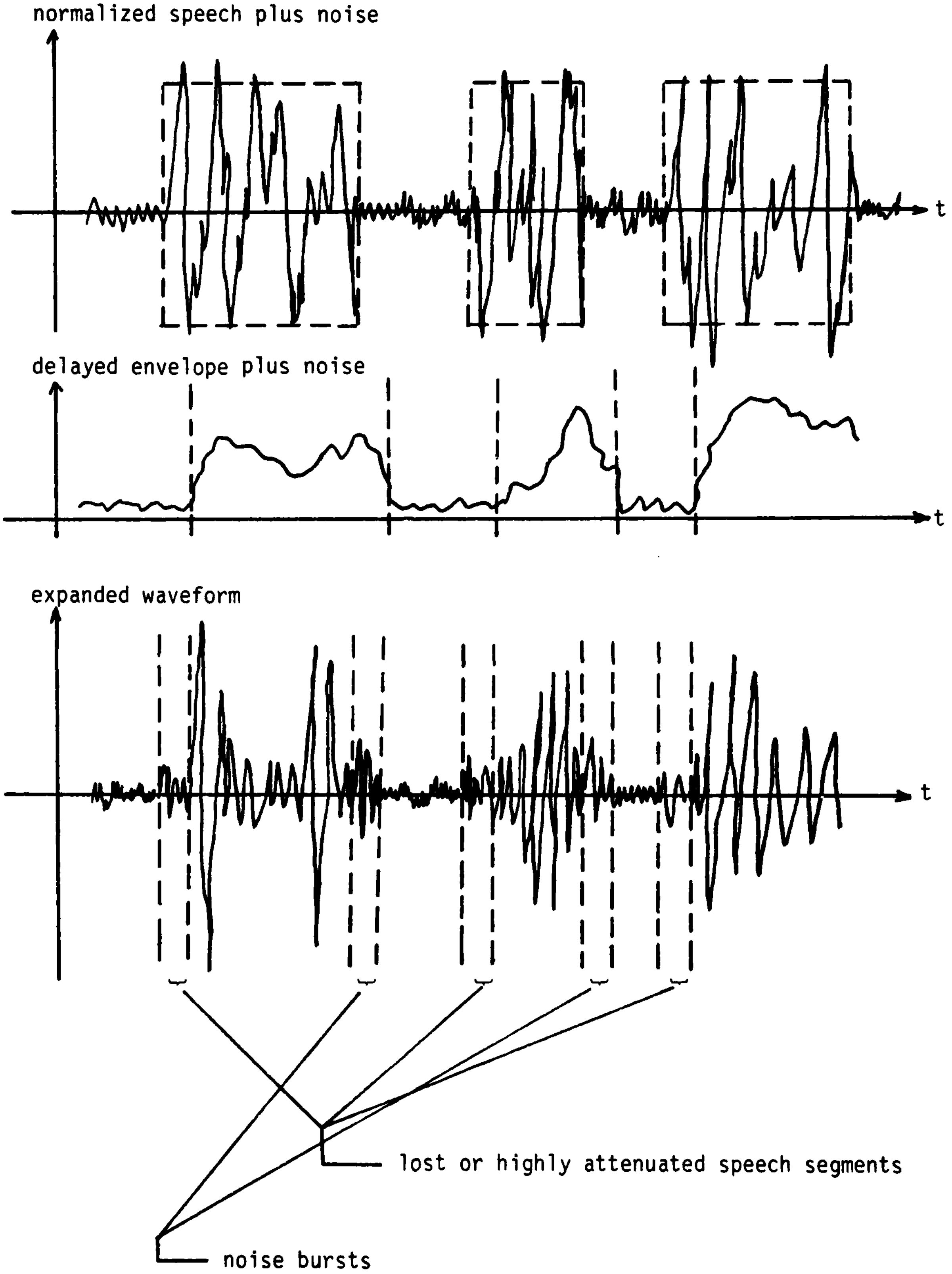

normalized speech plus noise
t
delayed envelope plus noise
t
expanded waveform
t
lost or highly attenuated speech segments
noise bursts

FIGURE 24 - ILLUSTRATION OF IMPROPER EN SPEECH AND ENVELOPE ALIGNMENT

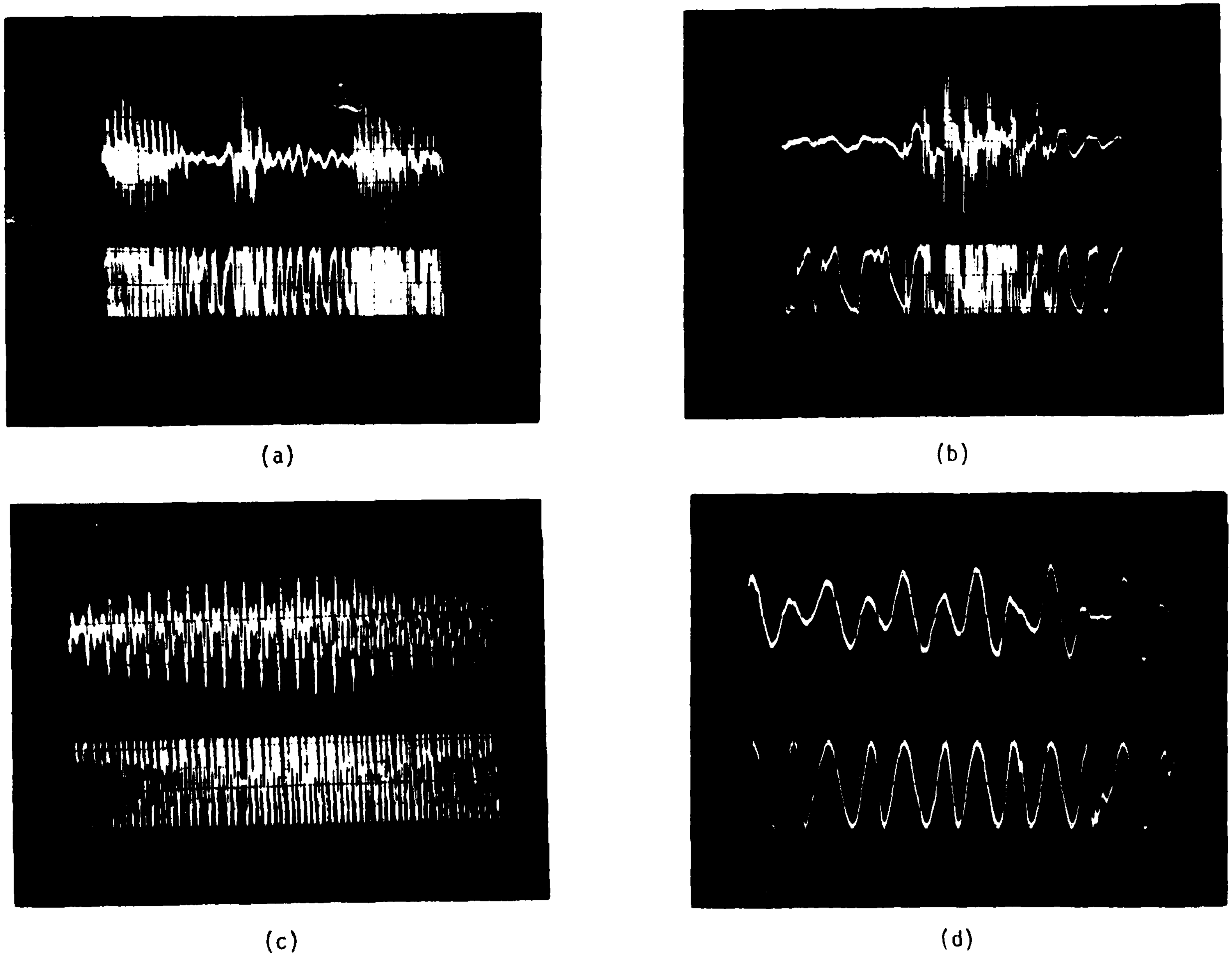
9.0 EXPERIMENTAL RESULTS

This section presents a summary of pertinent measurements made on the EN speech system. Temporal, spectral, and subjective listening observations are addressed. It is believed that many of the results represent an assessment of speech properties that are new or have not been revealed in the literature.

9.1 Envelope Normalized Speech Temporal Characteristics

Figure 25 shows four examples of speech segments before and after EN (with no EN waveform filtering). In the (a) portion, three speech bursts (akin to the form of Figure 1) are manifest, between which is some low-frequency noise (hum and other background components on the magnetic tape from which the speech sample was reproduced). The constant envelope nature of the EN speech waveform is clearly evident, as is the amplification of the intra-speech noise segments to the full EN output level (see the latter part of subsection 3.3 for discussion). Figure 25(b) shows a single burst representative of an unvoiced sound. Note the obvious frequency difference between the speech and noise segments. In the (c) portion of Figure 25 the vowel sound I (as in bit) is represented. Finally, Figure 25(d) indicates the variable-frequency sinusoidal nature of the EN signal (see subsection 3.3).

Figure 26, parts (a) and (b), show the voiced fricative G (as in get), where the lower EN waveforms are respectively unfiltered and filtered. (The lowpass filter characteristic is the interpolation LPF magnitude shown in Figure 15.) Such segments represent the extreme in terms of bandwidth expansion due to EN. Thus, the filtered EN waveform (Figure 26(b)) indicates the maximum extent that an ideal EN signal can be eroded by filtering. Clearly the change is not significant (see subsection 9.2 for additional results on EN signal filtering). For voiced speech segments, no discernable change in the original and EN temporal waveforms is evident. The (c) portion of Figure 26 shows original and EN signals where a small noise level precedes the speech burst. Again, note that the EN noise portion is amplified to the full

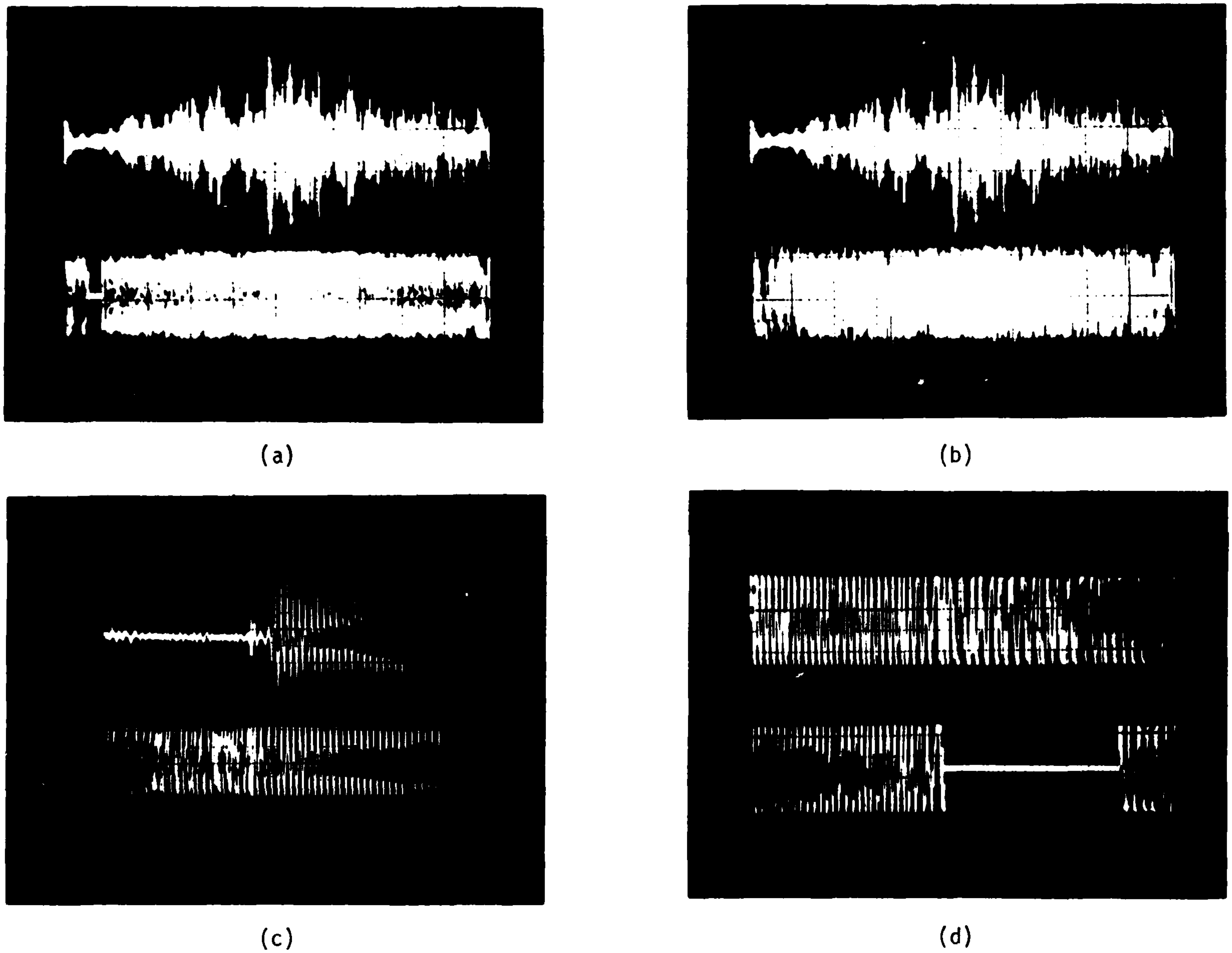

(a)

(b)

(c)

(d)

FIGURE 25 - ORIGINAL AND EN SPEECH WAVEFORMS


(a)

(b)

(c)

(d)

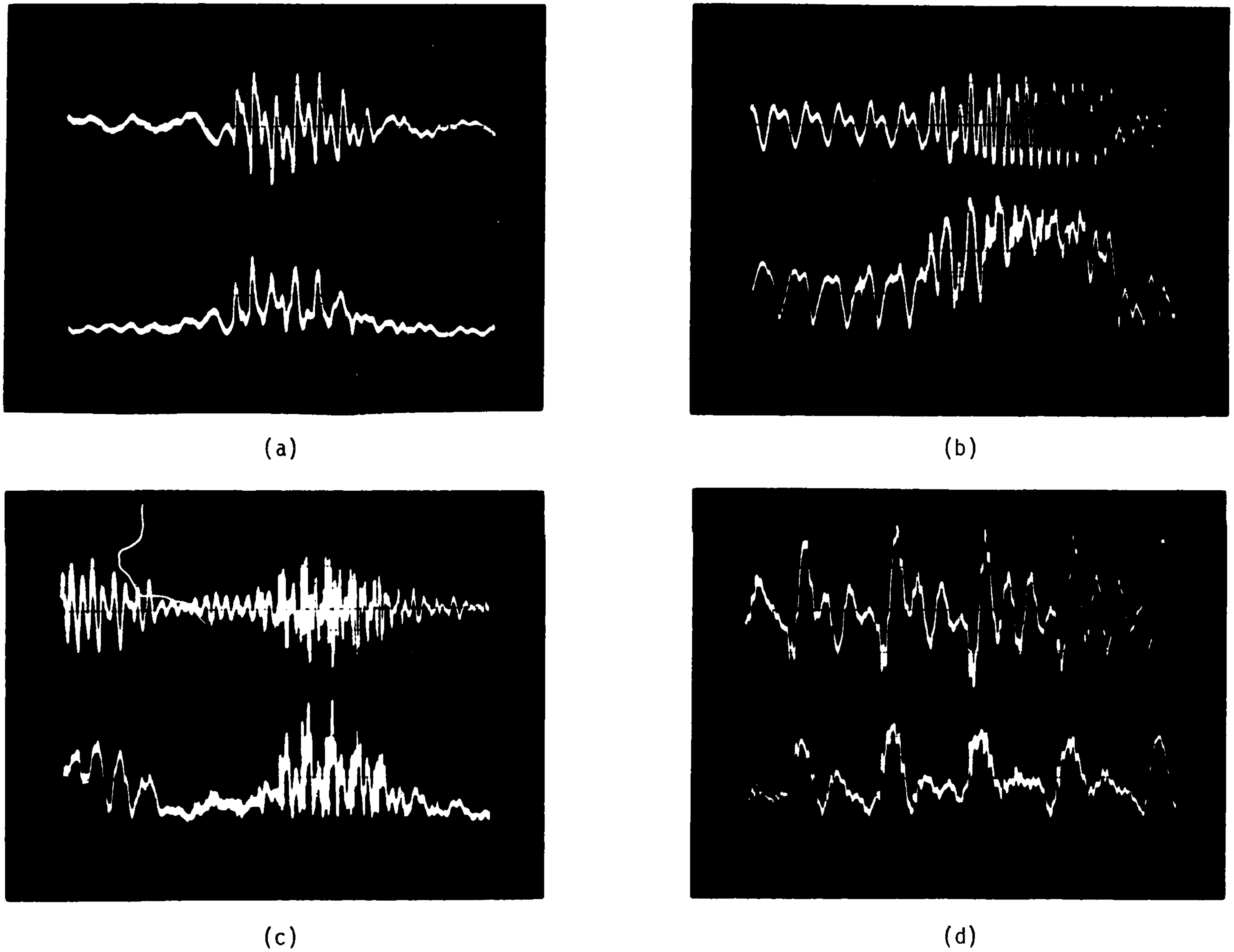
FIGURE 26 - FILTERED AND VOX'ED VERSIONS OF EN SPEECH

EN output level. Such noise can be eliminated using the VOX, and Figure 26(d) shows two EN waveforms that are respectively un-VOXed and VOXed. The VOXed segment represents a silence period between speech bursts.

Figure 27 shows four photographs wherein the upper waveforms are speech segments, and the lower traces are their true envelopes. Unvoiced speech segments produce rapid envelope variations, while voiced speech periods have a representative envelope embracing relatively lower frequency components (as in (b)). Envelope spectra are discussed in subsection 9.2. An interesting observation made from Figure 27 is that the speech envelope variations are more rapid than might be expected based on various data presented in the literature. Figure 28(a) shows a 0.5 second segment of speech with its corresponding envelope that clearly illustrates the presence of components much higher than the basic syllabic rate.

In the (b), (c) and (d) photographs of Figure 28 the upper traces show speech envelopes, and the lower traces are the envelopes subsequent to lowpass filtering. Single pole LPF -3 dB frequencies of 325, 100, and 20 Hz are represented in parts (a), (b), and (c) respectively. Note that the 20 Hz LPF virtually eliminates the high frequency components, leaving only the slow variation which depicts the syllabic envelope (usually called the speech envelope in the literature).

The importance of the speech envelope upper frequency components to the production of a high quality EN speech waveform is shown in Figure 29. The top portion of each photo is the same segment of original speech (vowel sound I). In the bottom traces of (b) and (c) the speech envelope was lowpass filtered at respectively 325 Hz and 100 Hz prior to its input to the EN speech divider, while the lower portion of part (a) illustrates EN without envelope filtering. Speech delay into the divider, to compensate for the delay introduced by the envelope LPF, was implemented using a clocked CCD delay line. Note the imperfect



(a)

(b)

(c)

(d)

FIGURE 27 - SPEECH AND CORRESPONDING ENVELOPE WAVEFORMS

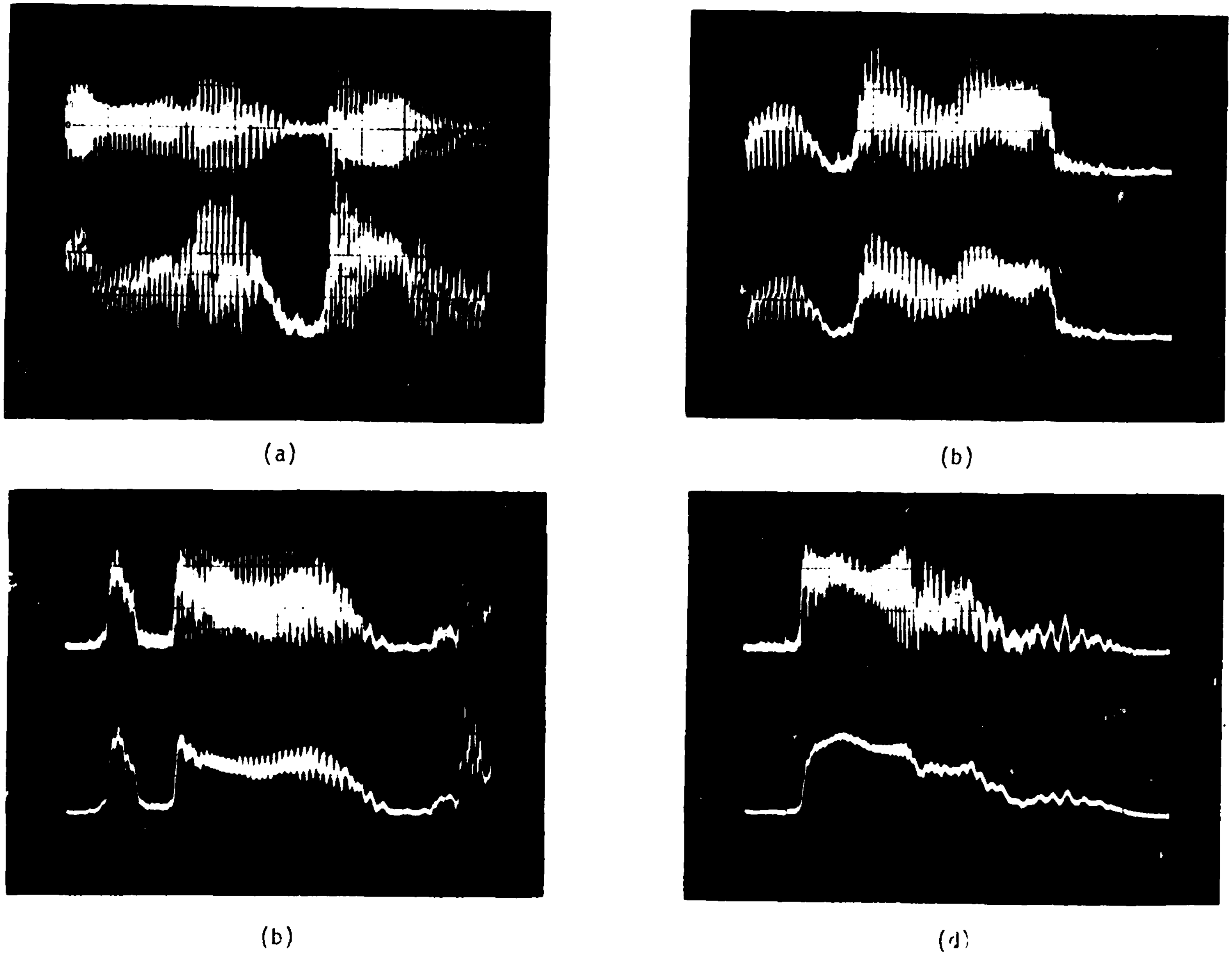
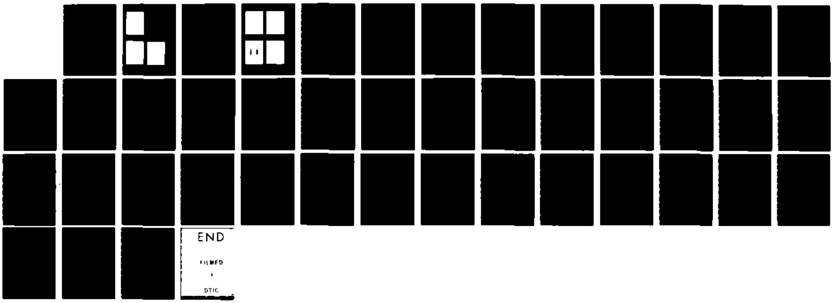
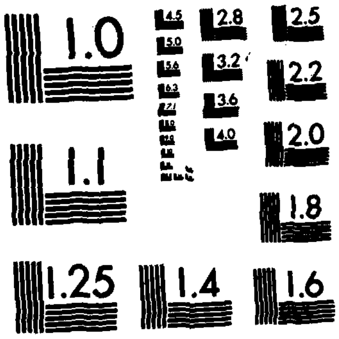


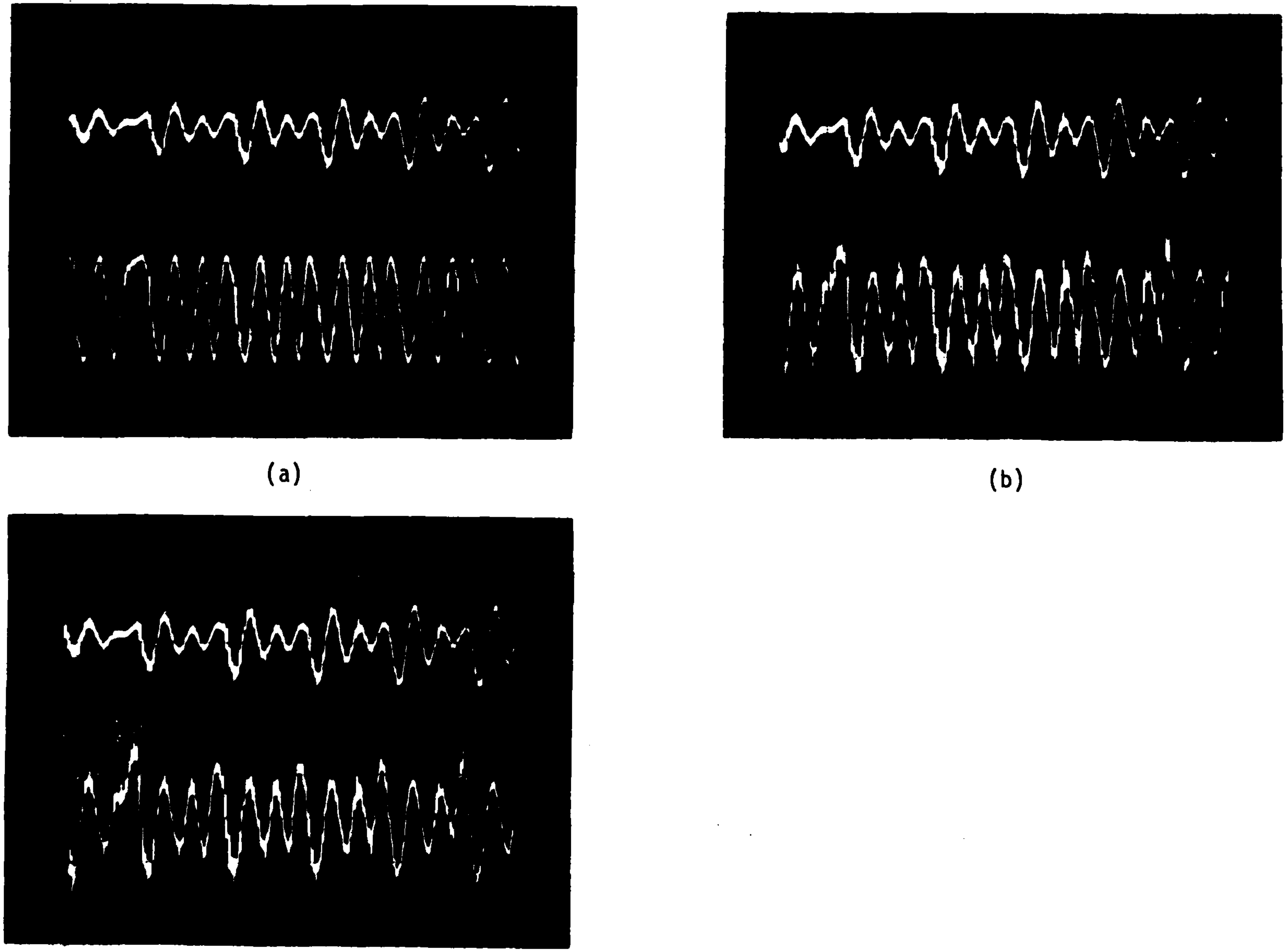
FIGURE 28 - EFFECTS OF LOWPASS FILTERING SPEECH ENVELOPES

AD-A124 225 SPEECH ENVELOPE NORMALIZATION A METHOD TO IMPROVE SNR 2/2
AND SUPPRESS NOISE I..(U) NATIONAL RESEARCH COUNCIL
WASHINGTON DC COMMITTEE ON VETERINA.. J C SPRINGETT
UNCLASSIFIED DEC 82 CECOM-82-J066-F DAAB07-82-C-J066 F/G 17/2 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



(a)

(b)

(c)

FIGURE 29 - EFFECT OF SPEECH ENVELOPE FILTERING PRIOR TO EN

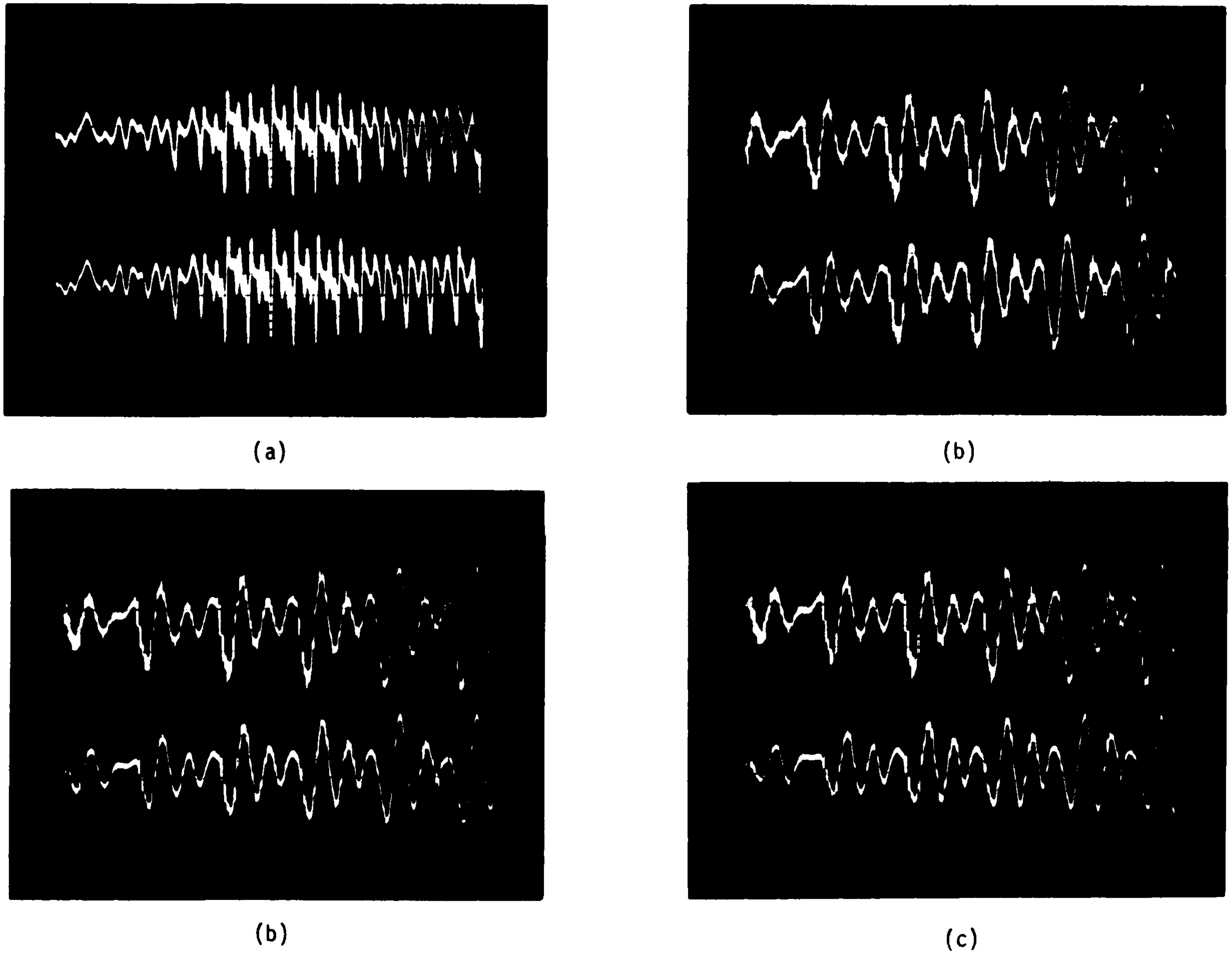
nature of the EN waveform in (b) even with the 325 Hz LPF, showing the importance of the speech envelope components above 325 Hz to the formation of a good EN signal. With the 100 Hz LPF in (c), the result is far from ideal. These experiments graphically demonstrate that speech envelope filtering into the EN divider should not be performed. Filtering the envelope into the expandor, however, is another matter.

Figure 30 presents four illustrations of EN speech expansion. The upper trace in each case is the original speech waveform, while the lower traces result in passing the speech through the EN compressor followed by expansion. Parts (a) and (b) represent perfect expansion in that there is no filtering of the speech envelope into the expandor. Note that the expanded waveforms cannot be discerned as any different from the originals. In parts (c) and (d) of Figure 30 the envelope has been filtered respectively to 325 Hz and 100 Hz. (Note that in (b), (c), and (d), the original speech is the same segment of the vowel sound I). The effects of envelope filtering are seen in the subtle differences between the original and expanded waveforms. Even with the 100 Hz LPF the distortion is not dramatic. Subjective tests (subsection 9.4) have indicated that the distortion introduced by the 325 Hz LPF is barely discernable, while with the 100 Hz LPF the distortion is fairly noticeable. Thus, the demonstration breadboard employs a lowpass bandwidth on the order of 200 Hz at the output of the linking pilot demodulator as a good compromise for minimizing expansion distortion and maximizing envelope SNR.

9.2 EN Speech and Speech Envelope Spectra

In this section the results of spectral measurements using the HP-3582A FFT spectrum analyzer are presented.

Although the envelope normalizer is designed to work with speech signals, its performance with broadband noise at the input reveals some of the characteristic behavior of EN and the implementing circuits. Figure 31 shows the spectrum of the noise before and after EN. The input noise spectrum is determined by the response of the AF132 LPF (see Figure 15).



(a)

(b)

(b)

(c)

FIGURE 30 - EFFECT OF SPEECH ENVELOPE FILTERING INTO EXPANDOR

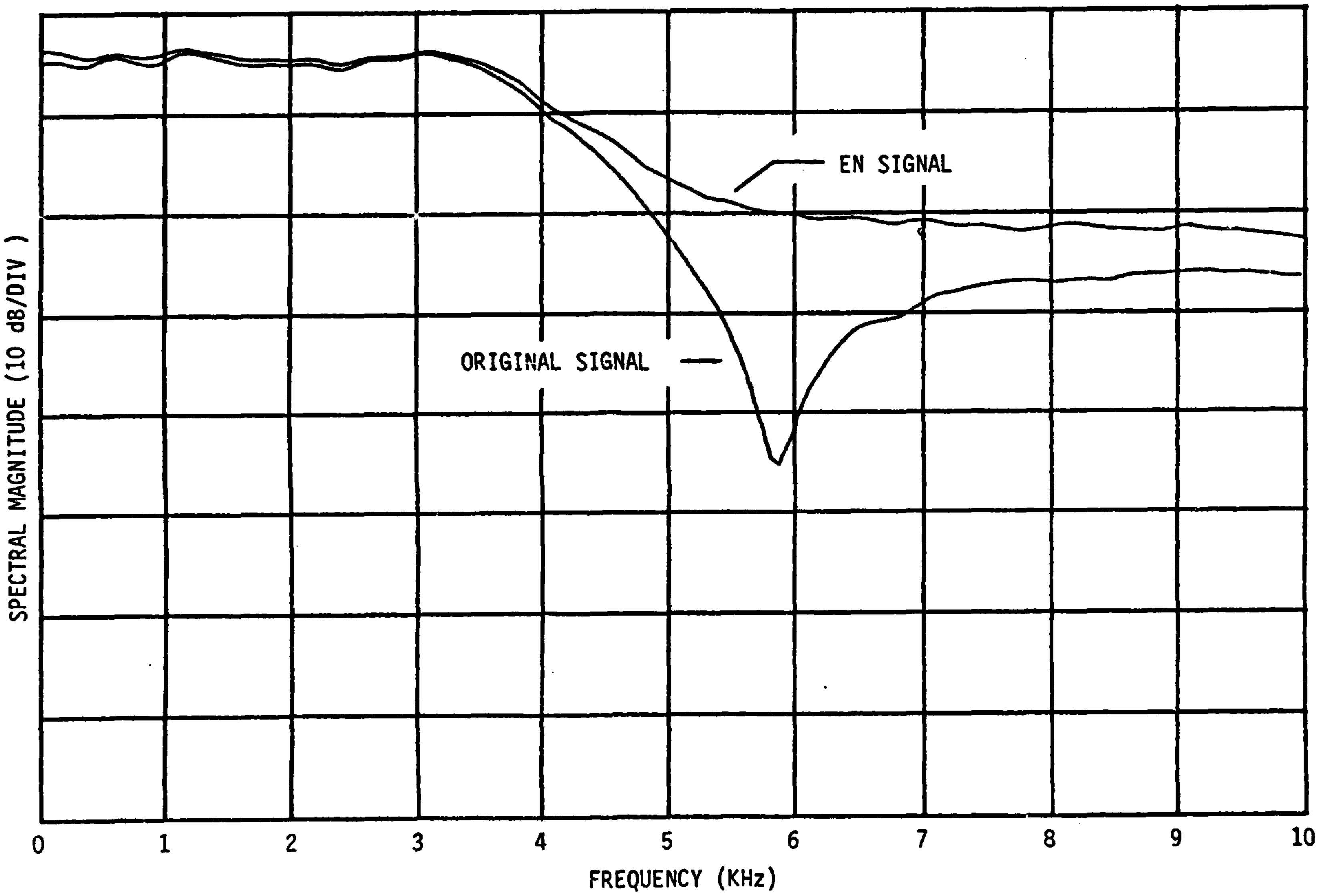


FIGURE 31 - SPECTRA OF LOWPASS NOISE BEFORE AND AFTER EN

It is seen that little in-band change of the spectrum results from EN, but that out-of-band the spectrum level is increased some 5 dB. It should be noted that the envelope of the noise also has a broad spectrum (unlike that of speech), so it is expected that the EN noise should show considerable spectrum expansion outside of the input spectrum effective cutoff. Figure 32 plots the EN noise after passing it through the interpolation LPF section of the AF132. As may be seen, the spectrum following the EN LPF is somewhat narrower than the original spectrum. This result stems from the interpolation LPF characteristic shown in Figure 15. What may be deduced from these measurements, plus the analytical results from subsection 3.3, is that the spectrum of speech following EN and the LPF should not appear wider than the input speech spectrum. Such is verified by speech spectrum measurements.

Figures 33 and 34 portray the average spectra for a male speaker, and depict the original speech signal, EN signal, and EN signal LPF results. In Figure 33 the creation of spectral components immediately outside of the input spectrum, due to EN, are in evidence. However, following the LPF of the EN speech signal, the original and EN spectra are virtually identical as illustrated by Figure 34. The average spectra consist of 256 individual spectra based on 25 ms sampling segments, the RMS average being taken. A special tape source of running speech in which lengthy pauses have been edited out was employed.

It is interesting to see what happens to the in-band spectra for certain basic sounds. Figure 35 shows the original and EN spectra ((a) and (b) respectively) for the vowel sound A (as in f<u>a</u>ther). The formant frequencies of this voiced sound are easily recognized (see Reference 19 for formant definition and a table of vowel formant frequencies). As may be seen, the second formant frequency is suppressed with respect to the first and third formants. This effect will tend to decrease the intelligibility of the sound if it is listened to in its EN version because the second formant frequency is more critical to articulation

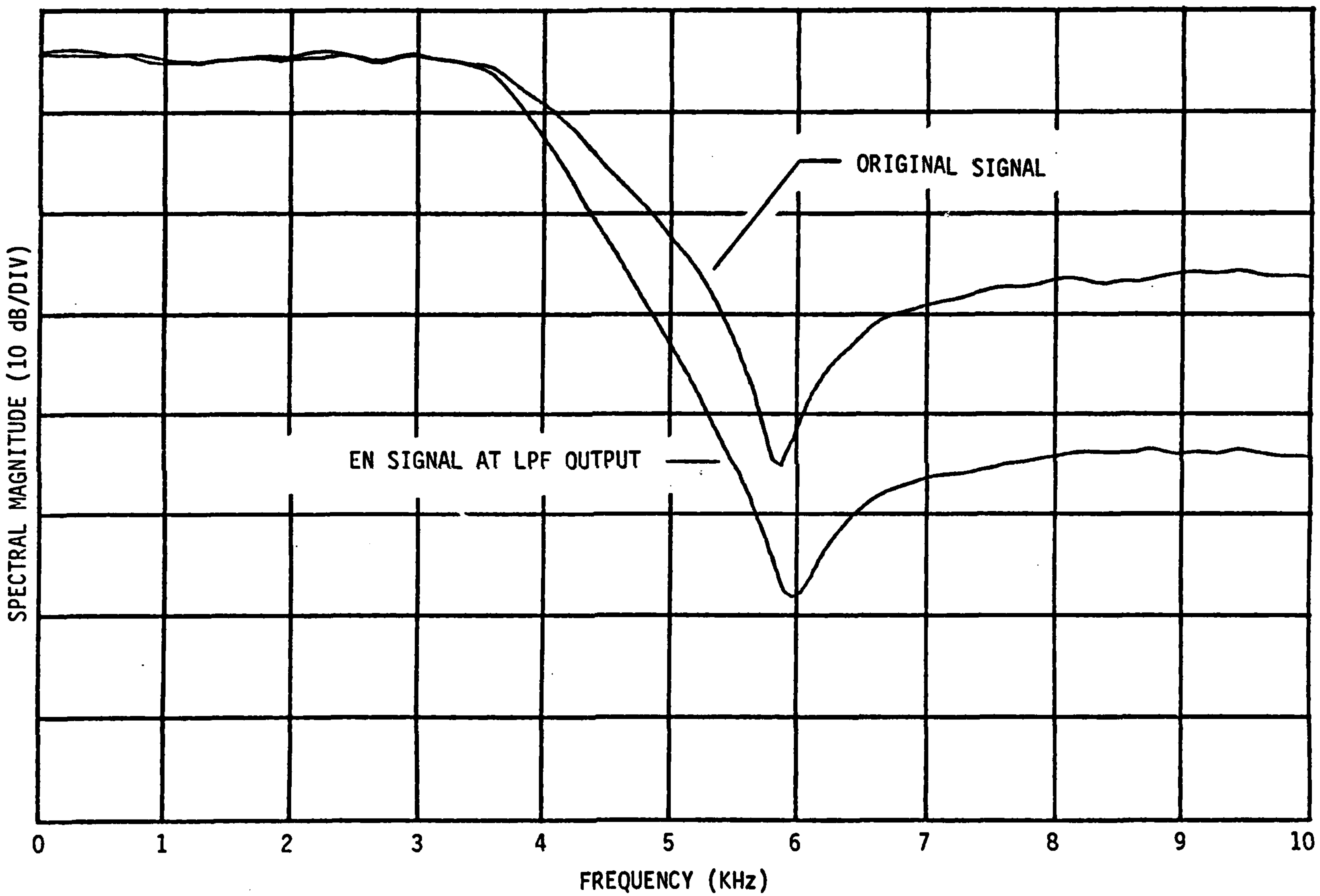


FIGURE 32 - SPECTRA OF LOWPASS NOISE BEFORE AND AFTER EN WITH FILTERING

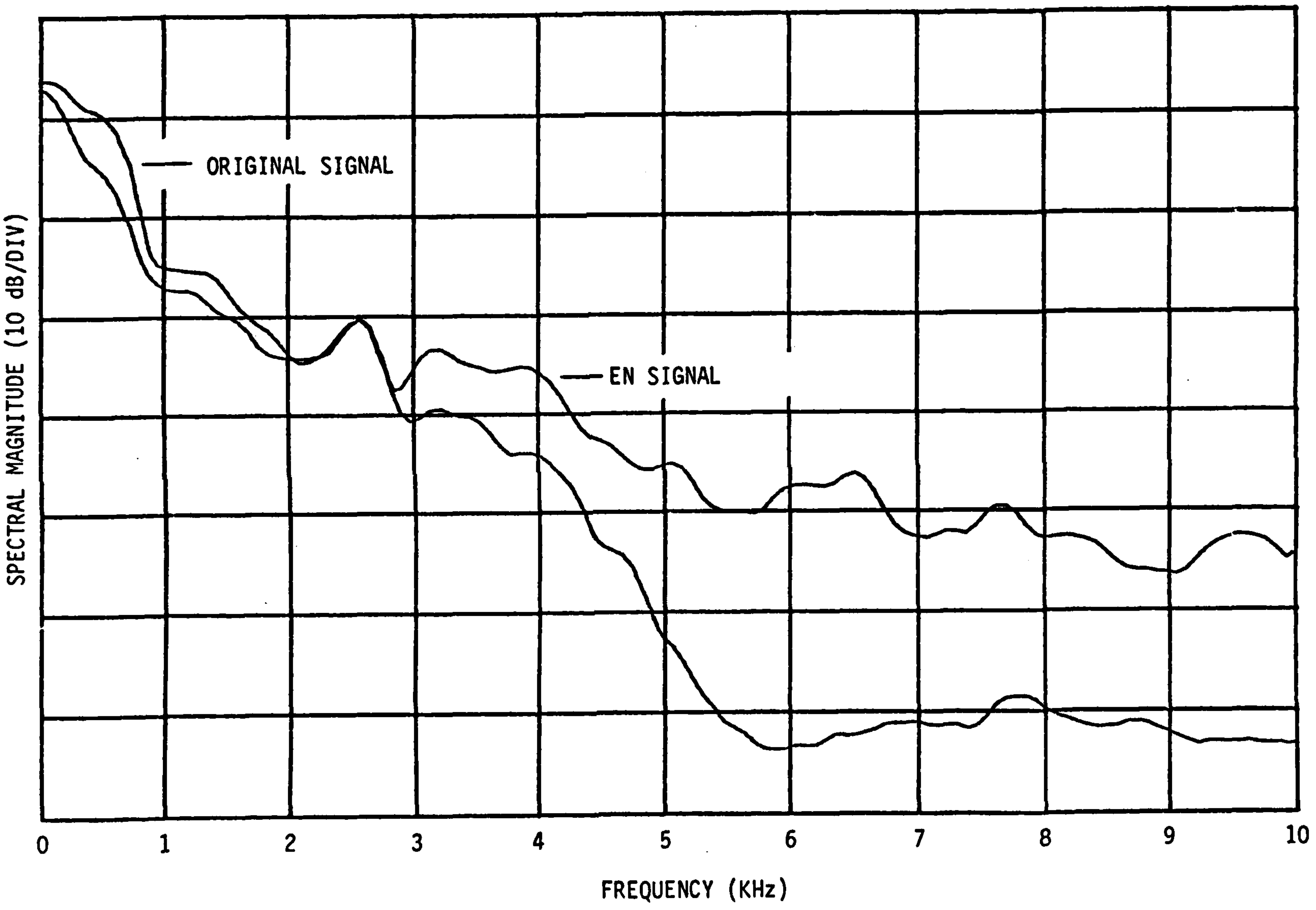


FIGURE 33 - AVERAGE SPECTRA OF MALE SPEAKER BEFORE AND AFTER EN

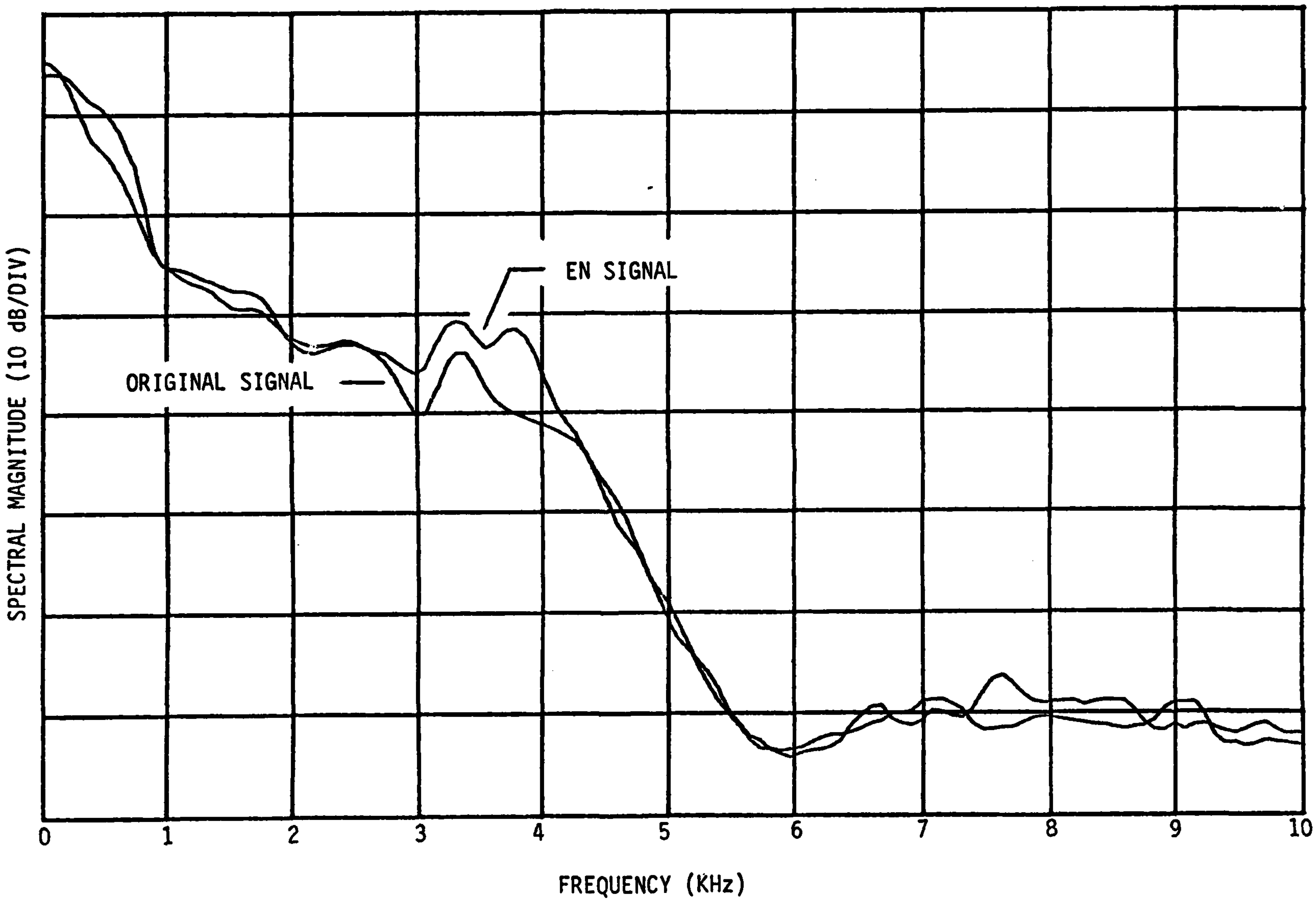
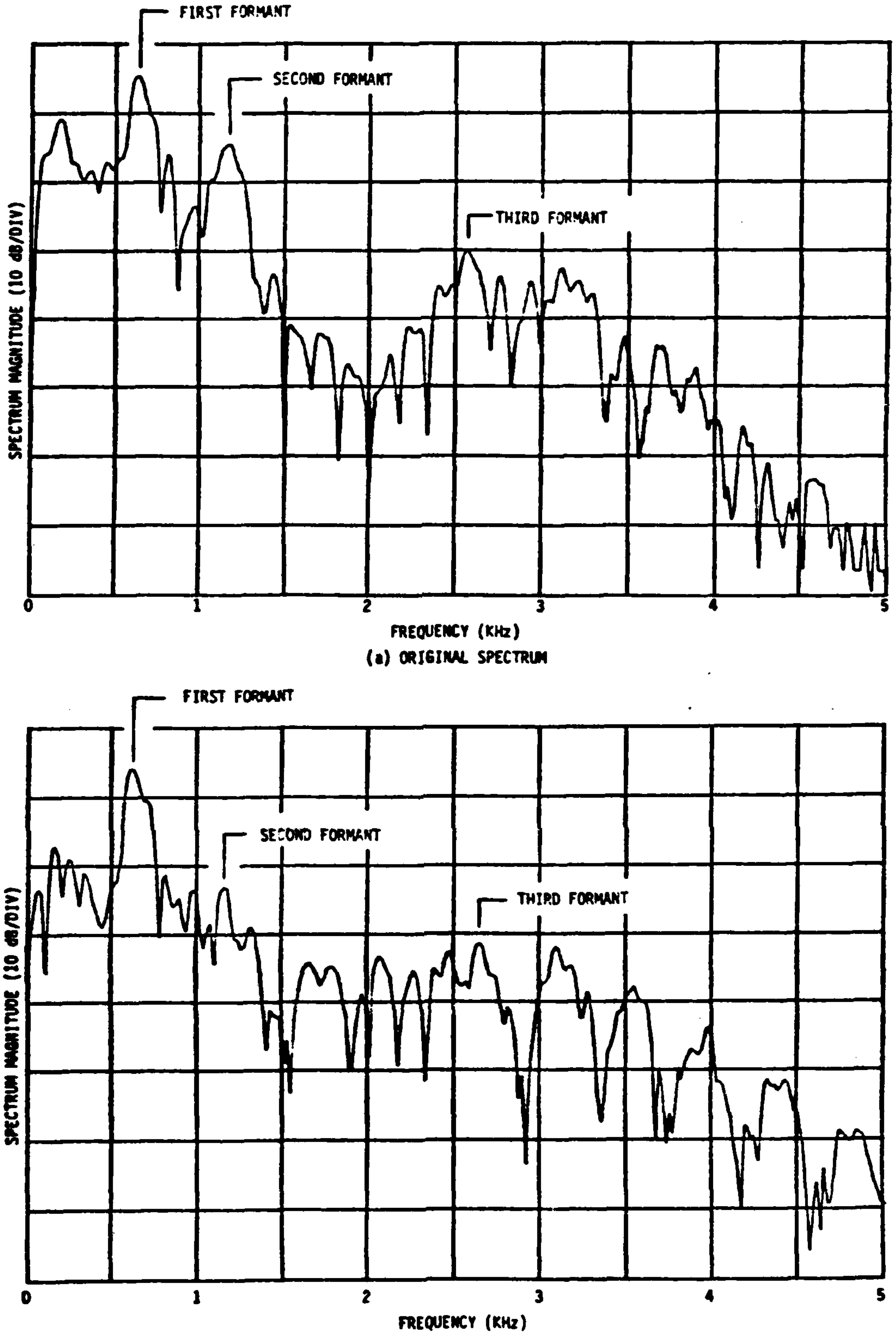


FIGURE 34 - AVERAGE SPECTRA OF MALE SPEAKER BEFORE AND AFTER EN WITH EN SIGNAL FILTERING



(a) ORIGINAL SPECTRUM

(b) EN SPECTRUM

FIGURE 35 - SPECTRA OF THE VOWEL SOUND A

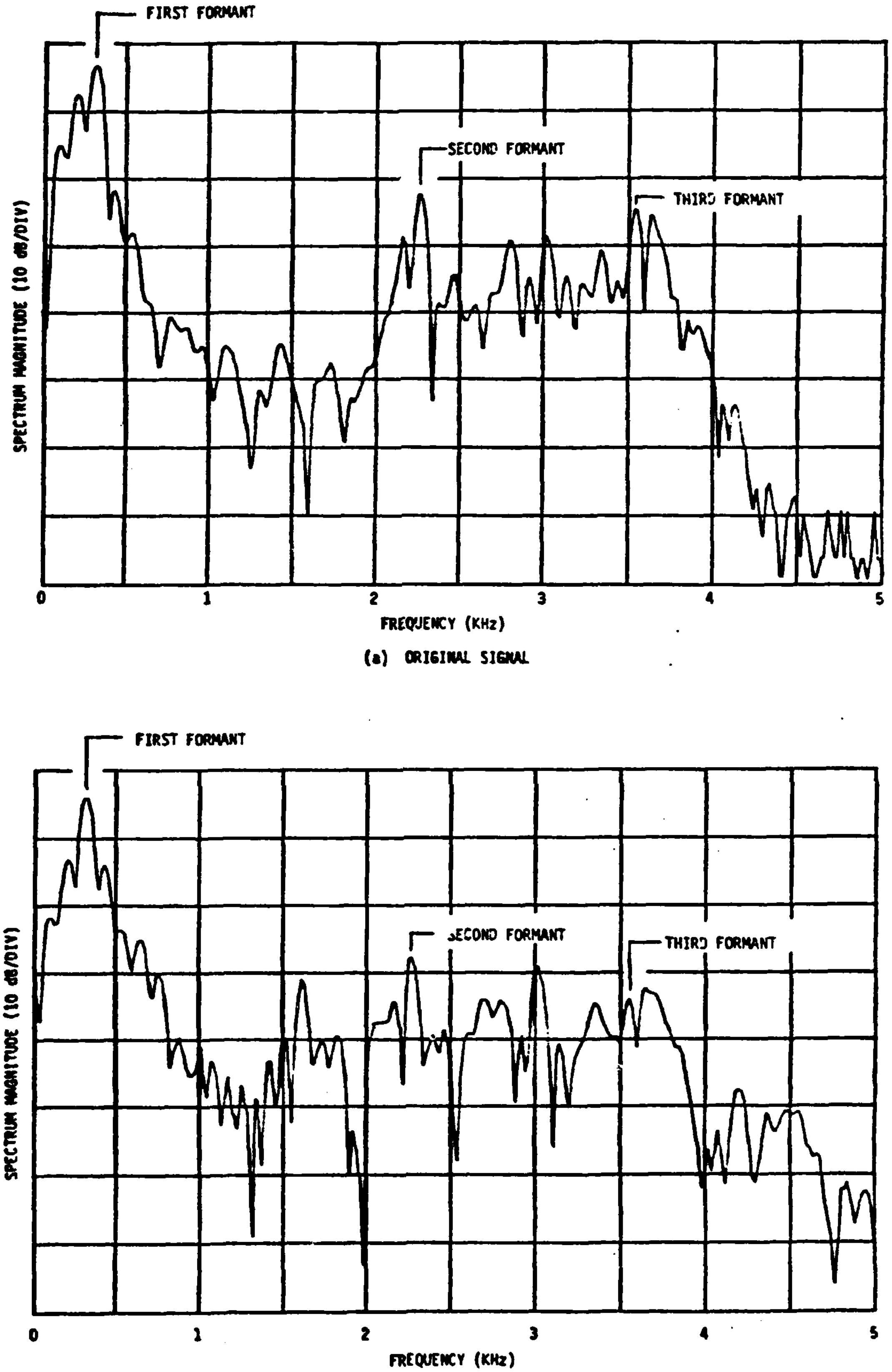
than the first formant.[24] Note in addition that the valley of the original spectrum around 2 KHz is increased in magnitude some 10 dB, while bandwidth expansion effects above 3 KHz are also evident.

The spectra of another vowel sound I (as in b<u>i</u>t) are portrayed as Figure 36. Again the observations made about the vowel sound A are repeated. Furthermore, two additional peaks or maxima at approximately 1.6 KHz and 3 KHz are evidenced. It is conjectured that the component at 1.6 KHz might represent an intermodulation product between the formant frequencies (perhaps $F_2 - 2F_1$). The peak at 3 KHz appears to be an amplification of an original signal component. Whatever the causes, these effects further degrade articulation when listening to the EN speech. A method for minimizing these latter phenomena is discussed in subsection 9.3.

As the last example of a basic sound, the spectra for the voiced stop G (as in <u>g</u>ut) are shown in Figure 37. It may be seen that there is little change of the original spectrum by the EN process. This is not too surprising since the sound is more noise-like than a vowel sound, and it has already been established that the in-band spectrum of noise is not significantly affected by EN.

Attention is now turned to the speech envelope spectrum. Figure 38 is the speech envelope spectrum corresponding to the male speaker average spectra of Figures 33 and 34. Note that the spectrum magnitude has a linear rather than logarithmic scale. It is seen that the spectrum decreases betwen 0 Hz and 20 Hz, and becomes minimum around 50 Hz. There is then a second maximum or lobe in the vicinity of 100 Hz. The spectral characteristics between 0 Hz and 50 Hz are definitely related to the syllabic rate of the speech and correspond to an envelope temporal record such as that shown in the lower trace of Figure 28(d). Figure 39 portrays the syllabic spectrum between 0 Hz and 25 Hz using a 2 dB per division magnitude scale. Note that the maximum occurs in the vicinity

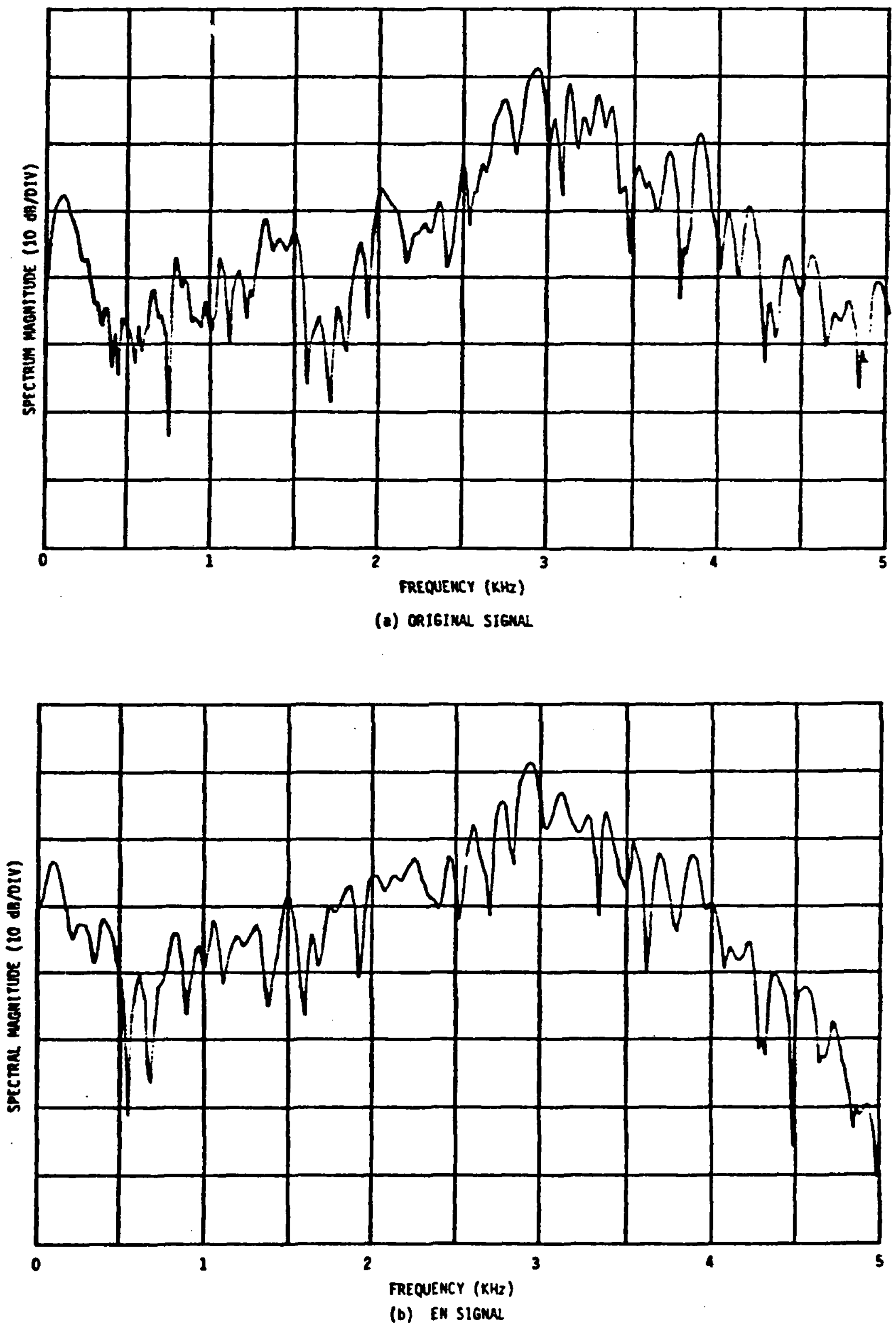
[24] Thomas, I. B., "The Second Formant and Speech Intelligibility," Proceedings of the National Electronics Conference," Vol. 23, 1967.



(a) ORIGINAL SIGNAL

(b) EN SIGNAL

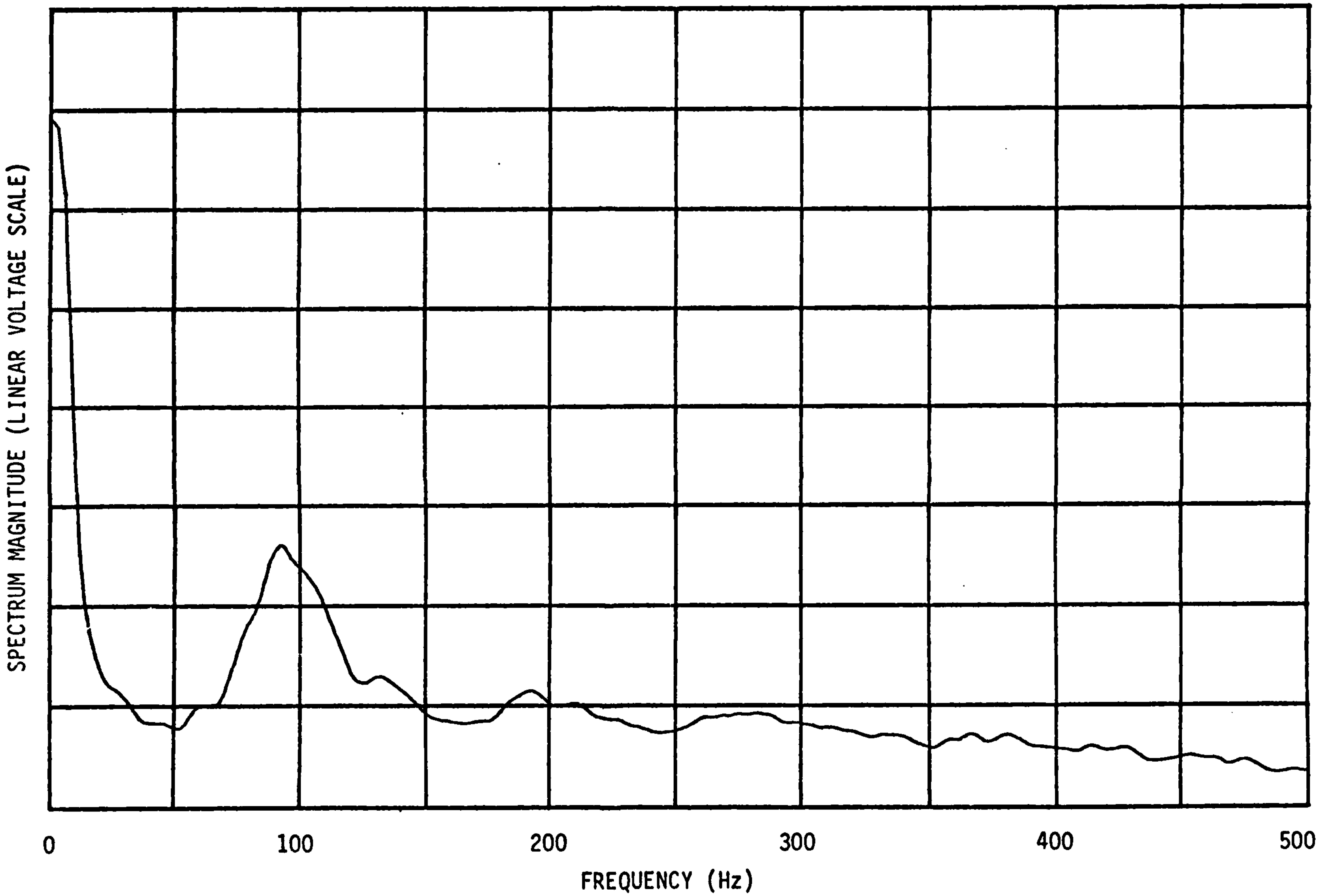
FIGURE 36 - SPECTRA OF THE VOWEL SOUND I



(a) ORIGINAL SIGNAL

(b) EN SIGNAL

FIGURE 37 - SPECTRA OF THE VOICED STOP G



AVERAGE 38 - AVERAGE ENVELOPE SPECTRUM OF MALE SPEAKER

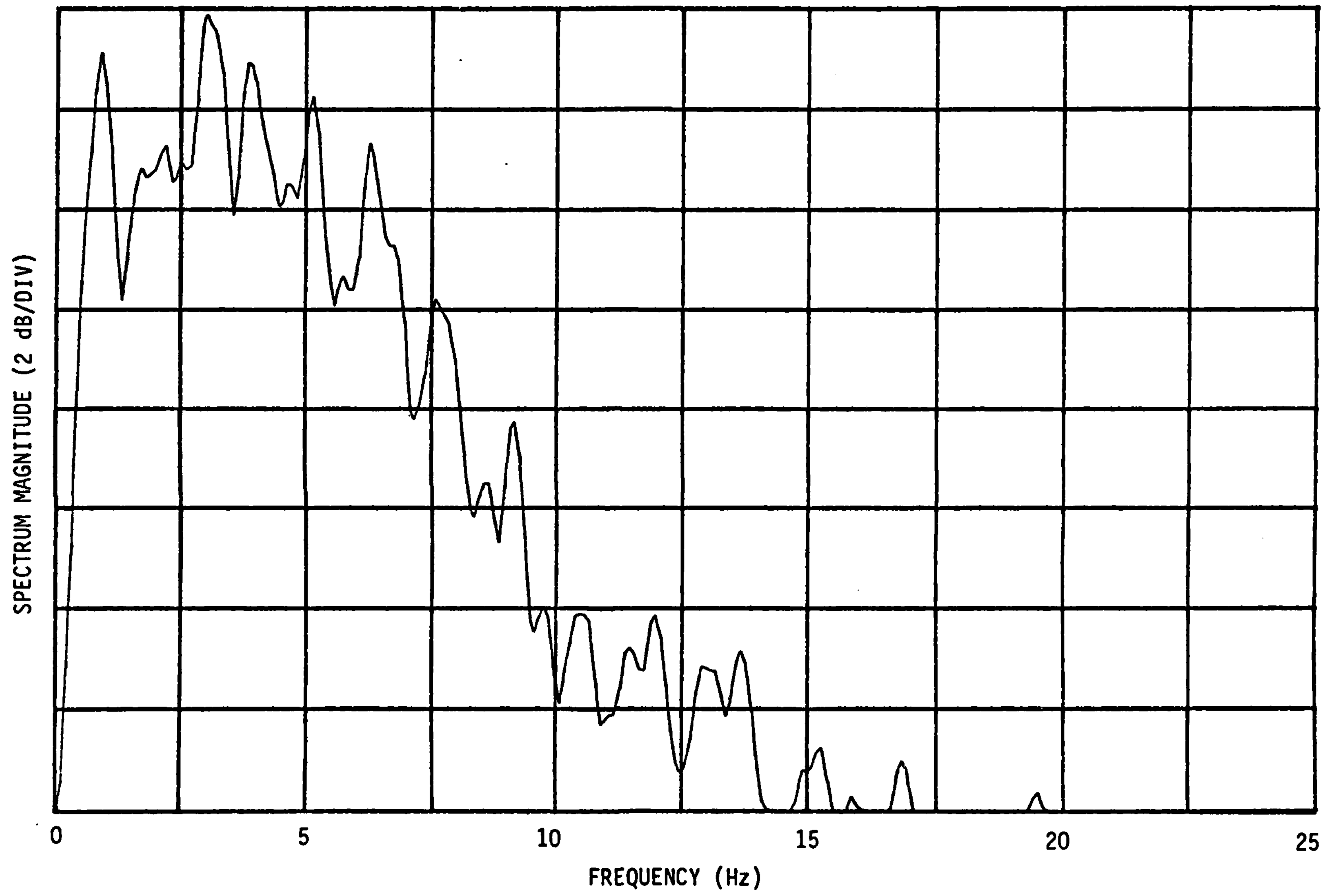


FIGURE 39 - AVERAGE ENVELOPE SPECTRUM OF MALE SPEAKER OVER PRINCIPAL SYLLABIC FREQUENCY RANGE

of a few Hz as reported in the literature. (The zero frequency component has been suppressed in Figure 39.)

The lobe around 100 Hz in Figure 38 is curious. At first it was suspected that its presence might be related to the harmonic generation problem that arises through the use of the LH0094 vector magnitude converter in the envelope formation circuits (see subsection 6.1). Further experimentation, however, dispelled this conjecture. First, additional speech envelopes were generated based upon different male speakers, four of which are plotted in Figure 40. Careful examination shows that the lobes for two of the speakers occur around 100 Hz, while for the other two speakers the maxima are around 140 Hz. Additionally, the amplitude of the lobes varies between the speakers even though the spectra below 50 Hz are essentially matched. The frequency location of the lobes appears to depend on the general pitch of the speaker's voice, with the higher frequency lobes occurring for the more treble talkers. With this observation, the lobe frequency should be generally higher for a female voice, which indeed it has been found to be as illustrated in Figure 41 where the maximum occurs in the neighborhood of 175 Hz.

Finally, in order to prove that the speech envelope lobe is peculiar to speech, envelope spectra were generated for non-vocal musical selections, the results for a symphonic orchestra and dixieland jazz band being presented respectively in Figures 42 and 43. As is seen, the speech characteristic lobes are not in evidence. It is interesting to note that the envelope for music decreases much like speech in the first 25 Hz. However, above 50 Hz, the music envelopes tend to have whiter or flatter spectra than those of speech.

Further research and experimentation disclosed the underlying nature of the lobe; it is related to the envelope of voiced sounds. When voiced sounds are produced, they arise due to a series of glottal air pulses being forced through the vocal tract[25] (which can be viewed as a

[25] Mathews, M. V., et.al., "Pitch Synchronous Analysis of Voiced Sounds," Journal, Acoustical Society of America, February 1961.

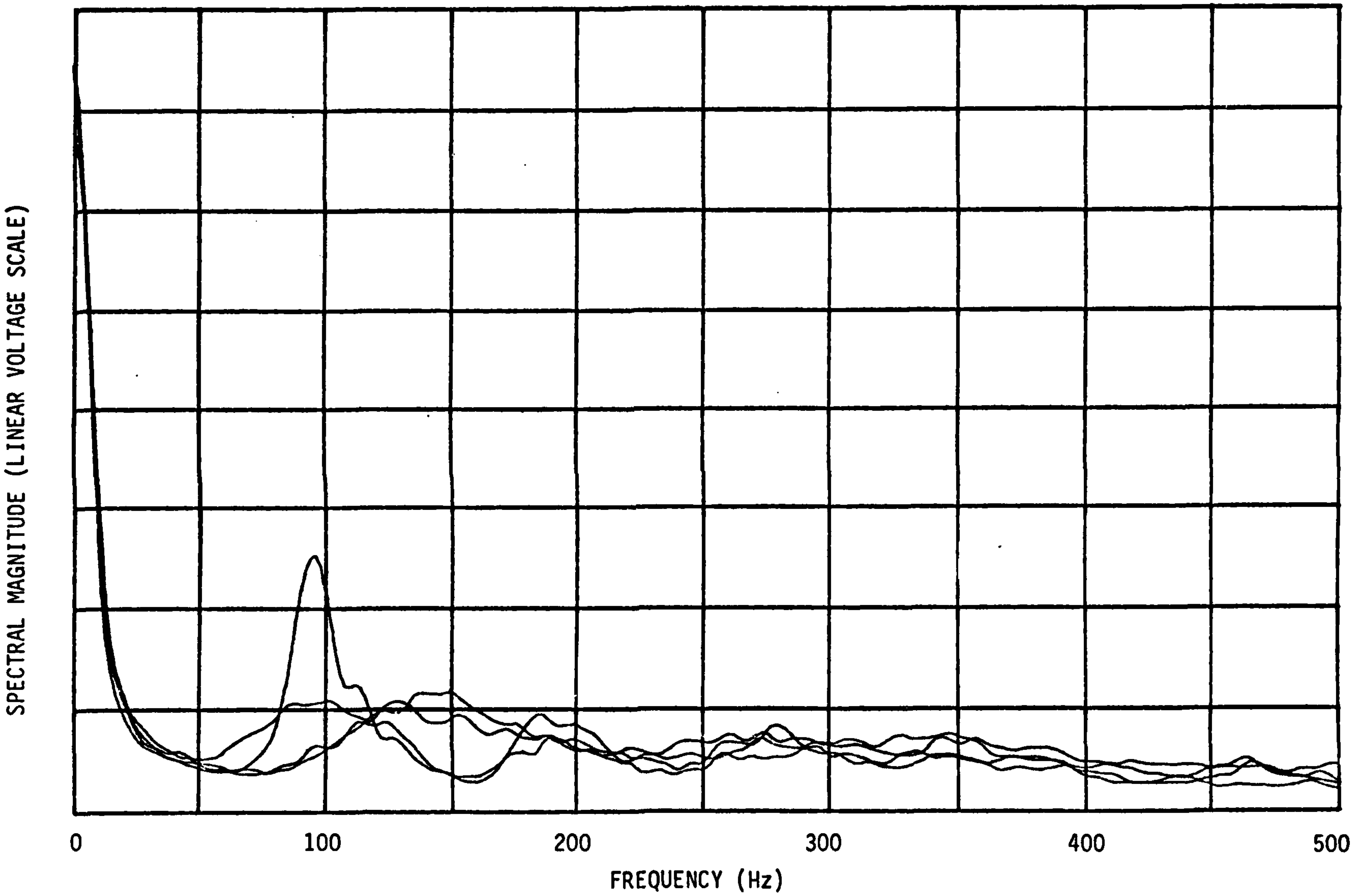


FIGURE 40 - AVERAGE ENVELOPE SPECTRA OF FOUR MALE SPEAKERS

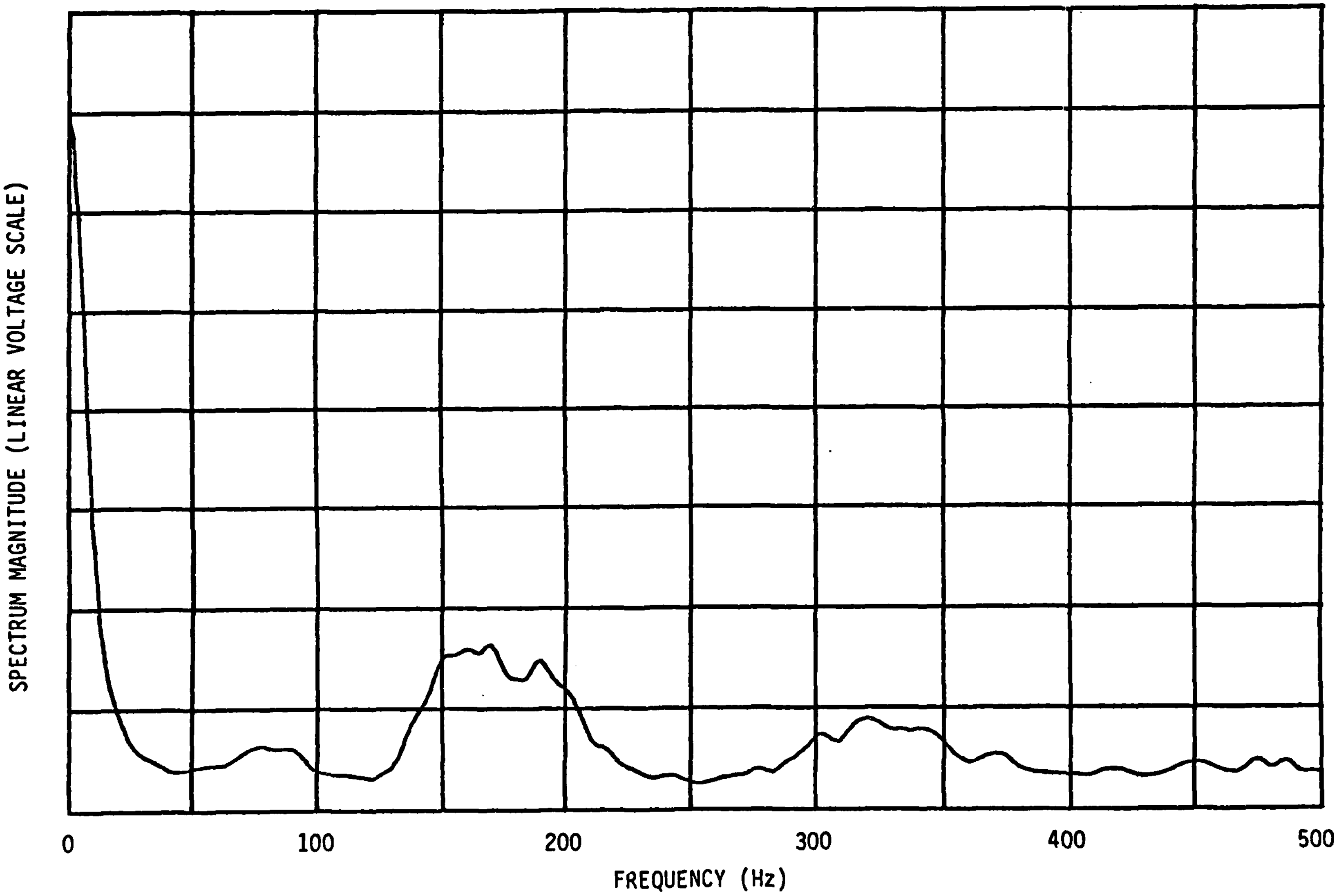


FIGURE 41 - AVERAGE ENVELOPE SPECTRUM OF FEMALE SPEAKER

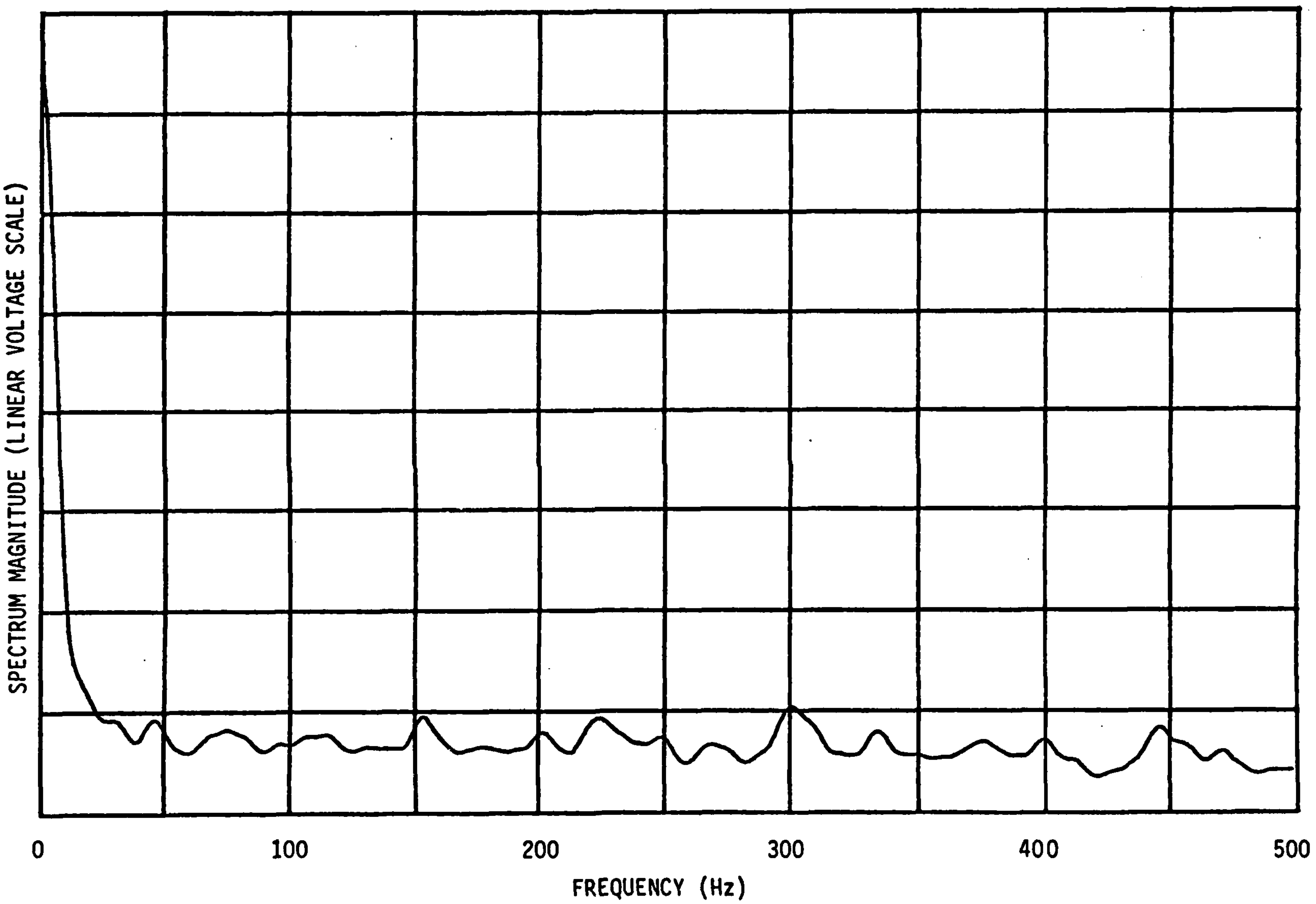


FIGURE 42 - ENVELOPE SPECTRUM OF A SYMPHONIC ORCHESTRA (MODERATE TEMPO)

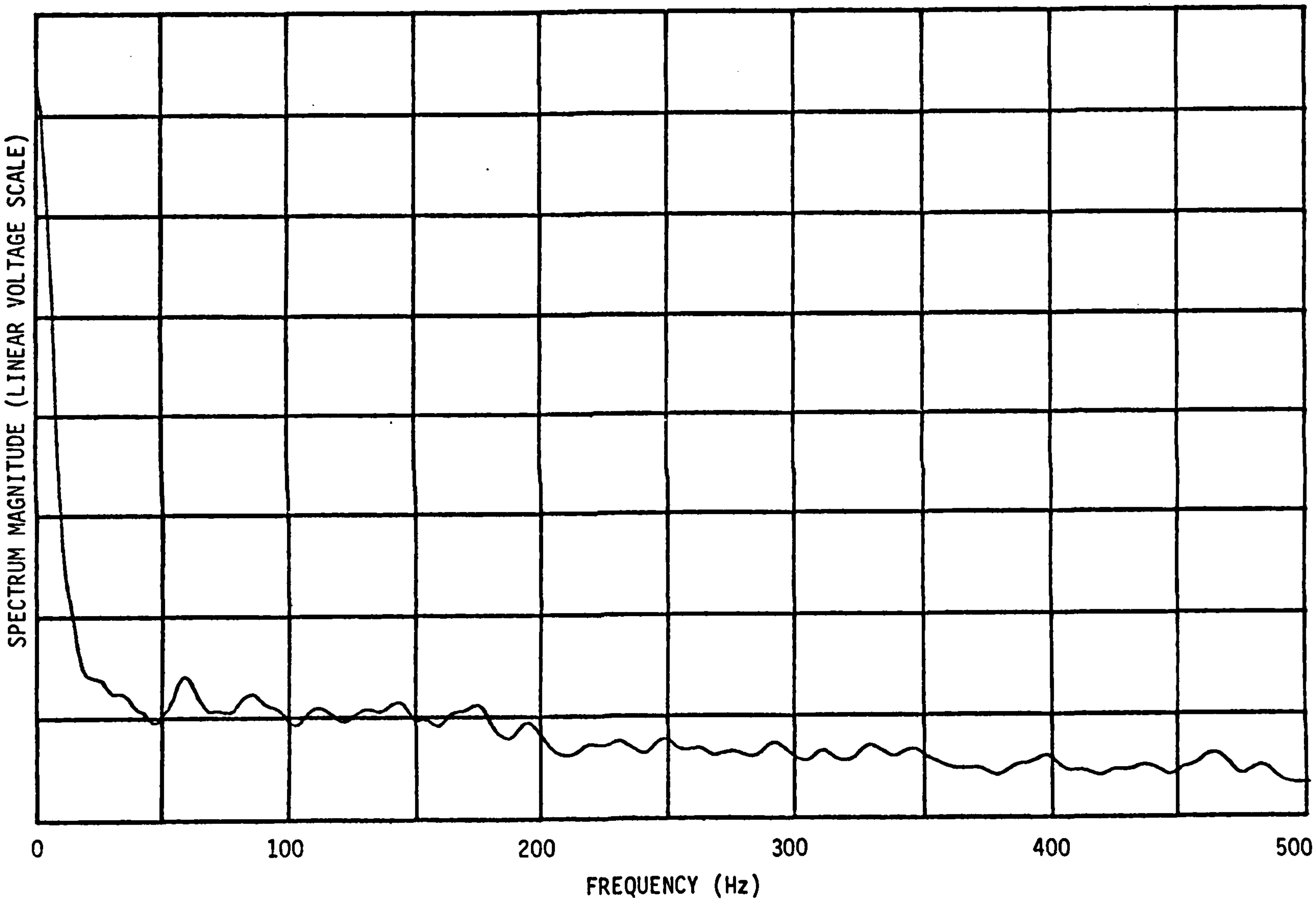


FIGURE 43 - ENVELOPE SPECTRUM OF DIXIELAND JAZZ BAND (FAST TEMPO)

time-varying filter). At conversational speech pitches, the glottal pulses are distinct, having a more or less fixed period, beginning and ending with discontinuities. As a result, each pulse "rings" the vocal tract filter, producing a disjoint series of decremented waves whose principal constituents are the formants. This process is illustrated in Figure 44.

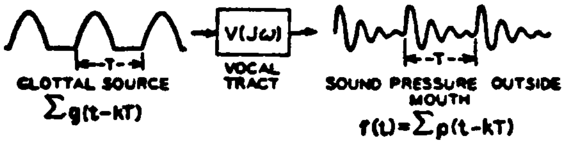


FIGURE 44 - VOICED SOUNDS MODEL

Experiments have shown that it is the repetitive envelope of the decremented waves that is responsible for the lobe in the average envelope spectra. When the pitch of a given voiced sound is changed by speaking inflections, emotions, etc., the glottal pulse rate changes, as does the vocal tract response somewhat, but the envelope of each of the decremented waves is reasonably invariant. Nor does the envelope shape fluctuate significantly between various types of voiced sounds. As a result, when a sufficient number of spectra for a variety of voiced utterances are averaged, the lobe, which represents a "smearing" of the various fundamental frequencies of the repetitive envelope shape, is produced. These fundamental frequencies, of course, are the corresponding glottal pulse rates, which may vary about some mean value by ± 30% to ±50% depending upon the individual talker. Also, because the voiced sound durations in conversational speech are reasonably short this also tends to diffuse what otherwise might appear as discrete spectral points. This also helps to explain why harmonic lobes are not prominent, although a second harmonic may be seen in Figure 41. Further, since the average envelope spectra are based on all types of sounds, both voiced and unvoiced, the weaker voiced envelope harmonics ultimately become masked by the averaging process.

By way of conclusion, it was shown in subsection 9.1 that envelope spectral components up to several hundred Hz are very important to the production of a quality EN speech signal. The preceding examinations of the speech envelope spectrum give further insight as to the reasons for the requirement.

9.3 Subjective Results Without Expansion

Subjective listening results are somewhat difficult to describe, but general qualitative assessments can be stated. Proper subjective evaluation requires that the reader make use of the EN demonstrator, or listen to a tape recording of the various EN speech signal conditions, and then form an opinion as to both good and bad properties. Because an opinion is required, the assessments made by a variety of listeners will vary to a degree (and be dependent upon training, past experience, etc.).

There are many practical communication situations wherein EN speech rather than expanded speech should be reproduced at the receiver. Such situations generally involve high-level background acoustical noise, so it becomes important to maximize the speech to acoustical noise ratio, especially for the unvoiced sounds which will be decreased in volume if expansion is employed. For this reason, then, the intelligibility and quality of EN speech is very important.

Subjective listening tests have been made on tne EN speech for both unfiltered and filtered conditions. Filtering makes little difference in perceptive quality. The two most prominent effects of EN speech are its "plosiveness" and a degree of raspiness. Certain speakers, especially those having higher pitched voices sound very clear and almost natural, while deep pitched voices sound somewhat mushy or muddled. In some cases it was noted that the effect of filtering the EN speech converted raspiness into mushiness.

It was discovered that the quality of the speech source has a direct bearing on the acceptability of the EN speech. Three basic sources were used during the evaluation program, microphones, magnetic tapes, and radio reception. It was found that using a radio as a speech source often led to poorer EN speech quality than when a microphone was employed. This is apparently due to broadcasting station tonal emphasis, distortion, and noise in the reception process. When using a radio as a speech source, the noise is often amplified to the full constant envelope level during speech pauses (as expected). With a noise cancelling microphone as the speech source, the ambient circuit noise level is quite low, and with the finite gain limit available from the divider circuit, noise amplification is generally below the constant envelope level. The tape player usually produces intermediate results, the biggest problem being the amplification of low frequency noise and hum during the speech pauses, which appears to degrade the speech-to-noise ratio.

Quantitative articulation measurements were not made during the current program because of the need for a trained listening jury, a proper listening facility, and the great amount of time needed to conduct the tests. Also, to be realistic and relevant to the assumed reason for listening to EN speech, such measurements should involve noisy acoustic situations. What was done, however, was to determine the general conditions which appear to foster a high degree of intelligibility.

It has already been mentioned that a treble type of voice seems to produce the clearest EN speech rendition. The absolute worst voice characteristic is one that growls. It seems that the former type produces the least in-band harmonic content, while the latter gives rise to many in-band distortion components. Envelope normalization of whispered speech produces excellent subjective results. The volume is constant, articulation is very high, and it sounds virtually natural. It should be noted that whispering is very noise-like in character, there being no voiced sounds as such. A conclusion drawn from these observations is that it is the alteration of the voiced sounds which prove most detrimental to EN speech articulation.

High pass filtering of the speech signal prior to EN will enhance the intelligibility of EN speech.[26] As was observed from the vowel sound spectrum plots in subsection 9.2, EN acts to suppress the second formant relative to the first formant. Thus, the purpose of the HPF preceding EN is to attenuate the first formant so that the EN process, in turn, produces first and second formants which have relative levels on the order of the unprocessed speech.

Niederjon tested a number of filter orders and cutoff frequencies for a type of amplitude compression similar to EN. (The degree of compression is not stated, but 8 ms. attack and release times were used for the syllabilic envelope estimation.) Best results based upon subjective testing were obtained for an HPF cutoff frequency of 2 KHz, and a roll-off of 6 dB/octive (an elementary RC section). This same filter has been tried with EN, resulting in a marked improvement of the EN speech quality. However, with the hope that even better results might be obtained, a graphic octive-band equalizer was employed. A large number of responses were tried. Optimum performance has been sensed when a 6 dB/octive preemphasis of the entire speech frequency band is used. Thus, this result is not significantly different from that reported by Niederjon, with the exception that the cutoff frequency occurs above the speech band. For the EN demonstrator this cutoff has been set at 8 KHz, and Figure 45 shows the measured response of the speech channel prior to EN when preemphais is employed.

[26] Niederjon, R. J., and J. H. Grotelueschen, "The Enhancement of Speech Intelligibility in High Noise Levels by High-Pass Filtering Followed by Rapid Amplitude Compression," IEEE Transactions on Acoustics, Speech, and Signal Processing, August 1976.

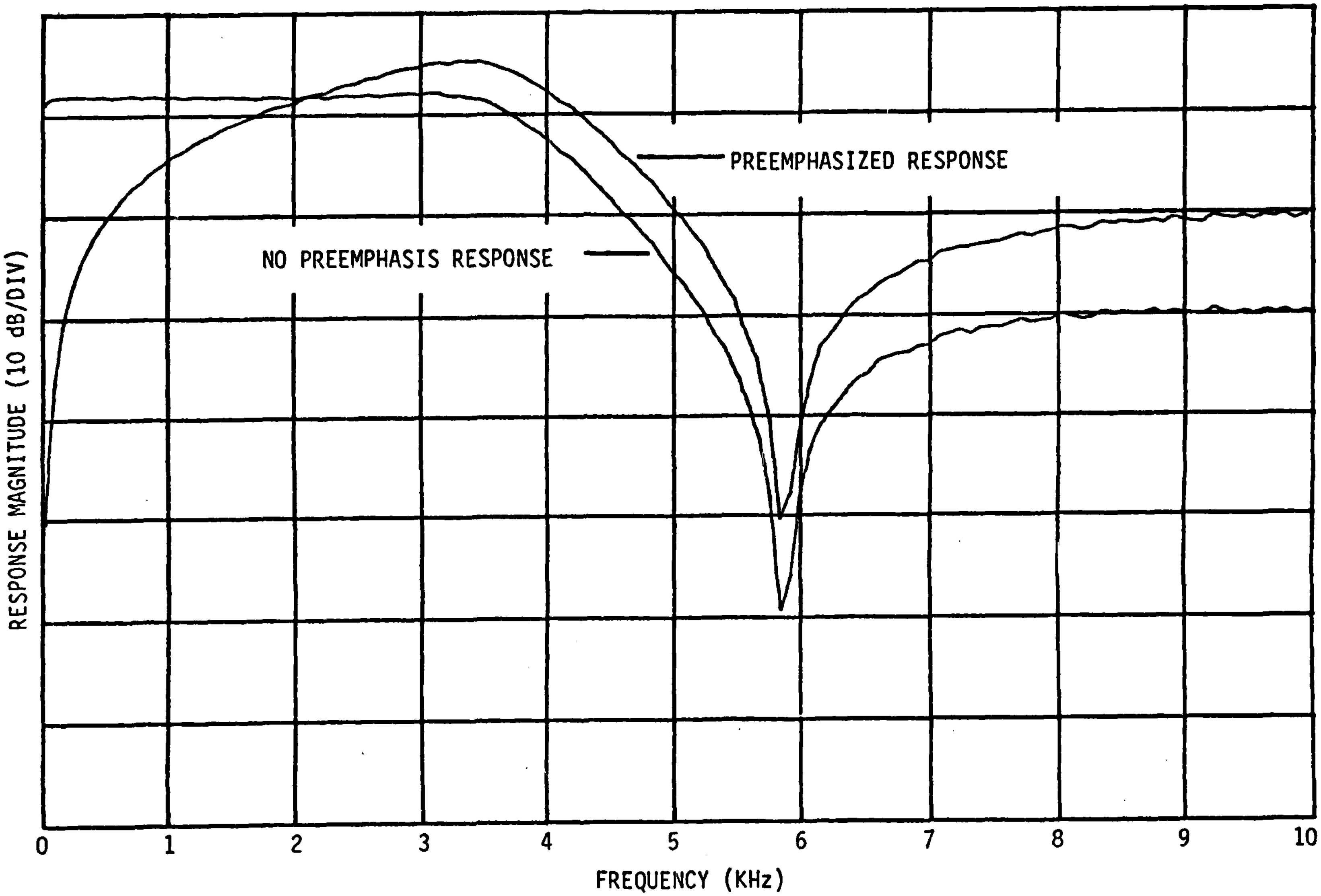


FIGURE 45 - FREQUENCY RESPONSE OF PREEMPHASIZED INPUT

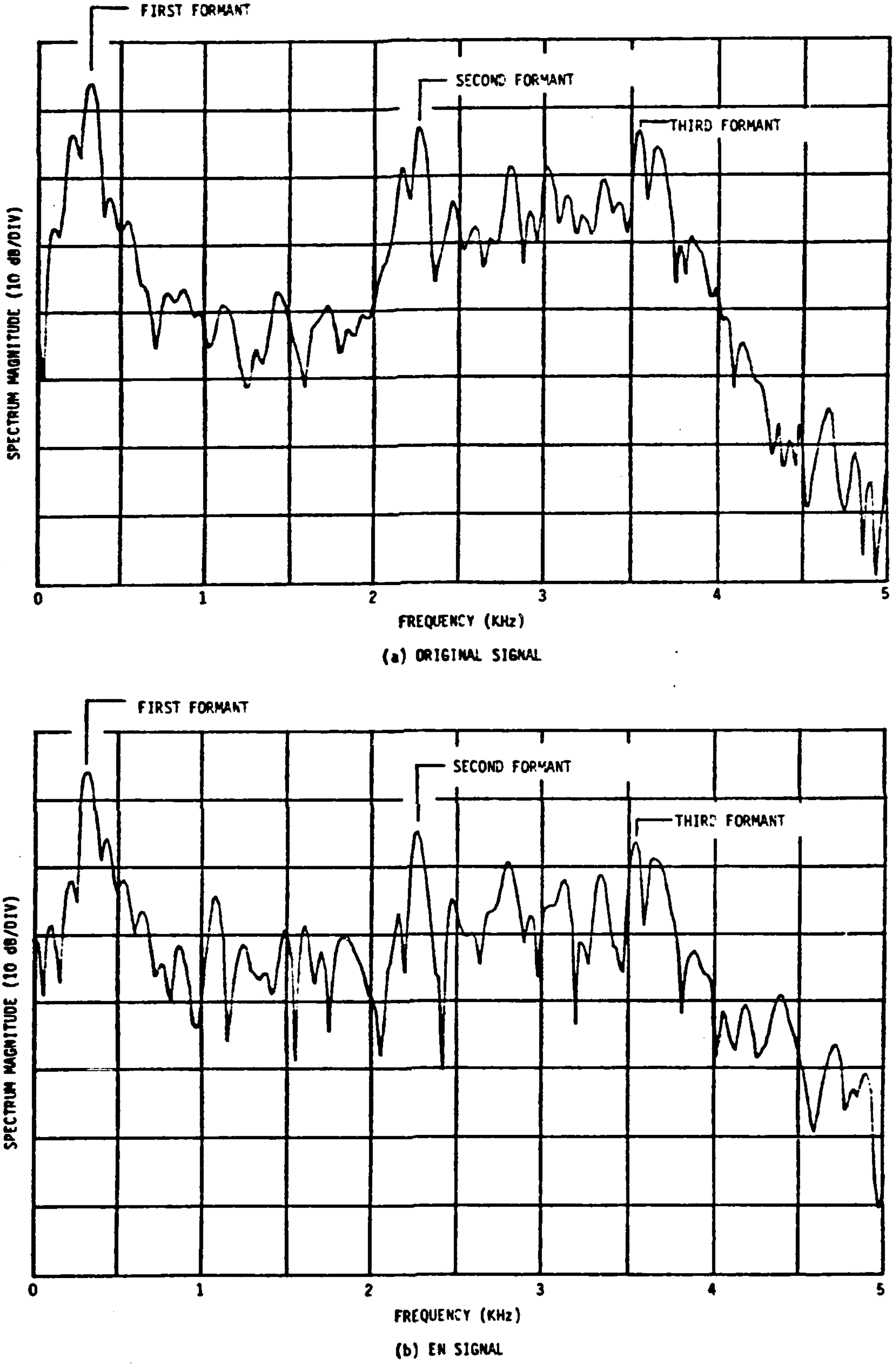
With the use of preemphasis, the subjective evaluation of listening to EN speech is that although there is still some raspiness present, the mushiness has been virtually eliminated, and intelligibility is very high. Probably the greatest detriment is the noise during speech pauses when the VOX is not employed. Subjective EN speech evaluation with the VOX in operation finds the EN speech to be far less objectionable than without the VOX. The seemingly poor speech-to-noise ratio due to noise amplification during pauses is much improved, and articulation is higher. VOXing, however, makes the EN speech appear even more plosive because the speech pause segments are totally quiet.

In summary, 6 dB per octive speech preemphasis produces a highly intelligible EN speech signal. That the EN process does not significantly alter preemphasized speech voiced sounds is illustrated in Figure 46, where the spectra for the vowel sound I are presented. These should be compared with Figure 36, where it can be seen that the degradations apparent in Figure 36(b) are not in evidence in Figure 46(b).

9.4 Subjective Results With Expansion

The subjective performance of the expandor has been evaluated for a number of conditions. First, listening tests were conducted with the transmitter VOX both in and out of operation. No significant differences in the speech were discerned, showing that the VOX function imposes no perceptible degradation on voice performance. The only VOX related effect that is noticeable occurs when any noise accompanying the original speech signal is high, and the VOX detector occasionally allows an EN noise burst to pass into the expandor. Since the envelope signal to the expandor will not be zero under this condition, a short burst of noise, albeit small, will be heard. If the VOX threshold is marginal, a number of successive bursts of this type may be heard as a "sputtering" sound. Such sounds may prove more annoying than the continuous noise signal, in which case the VOX should be disabled.

A second set of tests were made between filtered and unfiltered versions of the speech envelope used at the expandor. Filter bandwidths (-3 dB)



(a) ORIGINAL SIGNAL

(b) EN SIGNAL

FIGURE 46 - SPECTRA FOLLOWING PREEMPHASIS OF THE VOWEL SOUND I

of 500, 325, 150, and 100 Hz were tried. A CCD delay line was used on the EN signal to compensate for the envelope delay introduced by the LPF. Tests were also conducted without EN signal delay compensation. All comparisons were made with respect to the subjective performance obtained with an unfiltered envelope. With the 500 and 325 Hz bandwidth no expandor performance degradation was detected. When the 150 Hz LPF was used, some "fuzziness" in the expanded speech was evident, but not especially objectionable. For the 100 Hz LPF the "fuzziness" becomes somewhat annoying, but articulation is still high. No difference in performance with and without EN signal delay compensation was discerned except for the 100 Hz LPF, where the expanded speech, when the delay was removed, appeared to be slightly plosive, and the "fuzziness" was more apparent. (See subsection 9.1, Figure 30 for the temporal effects of speech envelope filtering into the expandor.)

Expandor performance with additive noise at the expandor input has also been evaluated. The expected subjective SNR improvement is very evident. This improvement does not appear to be at all changed by envelope filtering to the expandor. With sufficient noise present (speech SNR < 25 dB) even the "fuzziness" due to envelope filtering is effectively masked.

A central question was whether the subjective SNR improvement due to expansion could be measured. Since jury type testing could not be conducted, some other method was sought. It should be recalled from the discussion in subsection 4.2 that the improvement under consideration results from noise quieting during speech pauses. Thus, one possible way of measuring subjective improvement is shown in Figure 47. A single noise source is used for each channel. The channel with the expandor amplifies the noise by a gain G, while the envelope to the multiplier is normalized to its own standard deviation σ_e (thus it is a unit-power reference). The envelope signal e(t) is derived from a good noise-free segment of running speech. To make the measurement,

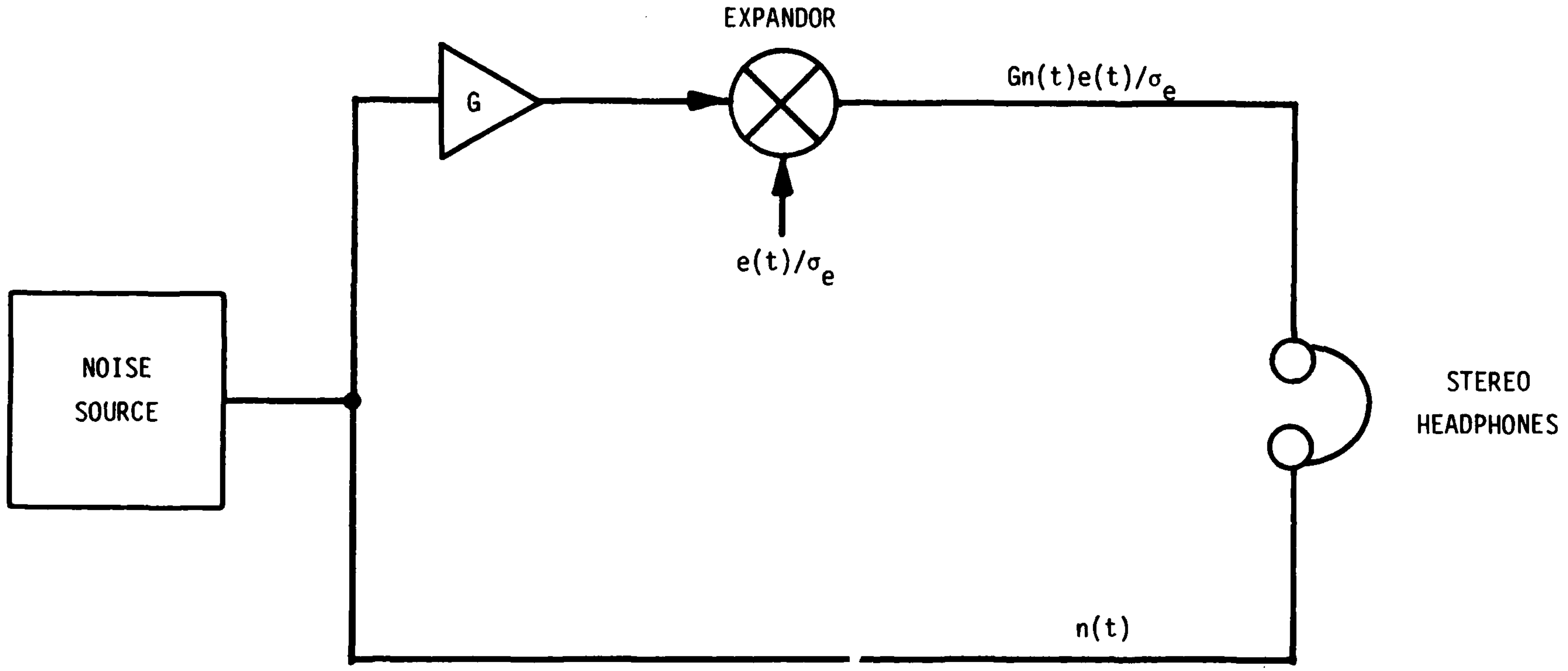


FIGURE 47 - MEASUREMENT OF SUBJECTIVE NOISE IMPROVEMENT

both channels are listened to simultaneously on stereo headphones, and G is adjusted until equal noise levels are sensed. The subjective improvement in dB is then

$$\text{Subjective Improvement} = 20 \log (G). \qquad (79)$$

However, obtaining a sense of balance between the steady noise and the envelope modulated noise was found to be quite difficult. The speech source employed was a radio tuned to a talk show. It was quickly discovered that the talking rate had a profound effect on the result. With "slow" talkers it was very hard to obtain a sense of balance as the ear tends to follow the speaking vs. silence intervals, rather than averaging across them. For "fast" talkers a sense of balance was obtained, but the results had a rather large variance from measurement to measurement. Improvements of between 4 dB and 10 dB were obtained, but it was finally concluded that the method is incapable of producing sufficiently consistent results as to be valid.

No additional techniques for measuring subjective improvement were devised. Therefore, quantitative results could not be obtained.

Finally, a few comments about expandor performance with envelope linking are in order. When there is no talking (i.e., $v(t) = 0$ and $e(t) = 0$), it is seen from eqn. (66) that the expandor output will consist of $n_1(t)n_2(t)$. The bandwidth of $n_2(t)$ is on the order of 200 Hz, while the bandwidth of $n_1(t)$ is about 4 KHz. As a result, $n_2(t)$ may be viewed as a relatively slowly varying modulation of $n_1(t)$. Listening to $n_1(t)n_2(t)$ sounds like a "crackling" noise rather than a "hiss."

To minimize this problem, a diode is placed in series with the envelope demodulator output to the expandor input. This has two effects; it functions to half-wave rectify $n_2(t)$, and acts to provide a small-voltage dead-zone. The latter operation effectively prevents $n_2(t)$ from reaching the expandor, thus, the crackling noise is virtually eliminated.

When speech is present, and the noise level is high so that the envelope demodulator is near threshold, a crackling noise can again be heard. However, if the envelope demodulator is above threshold, there is a little crackling noise, and the expandor operates in a near perfect fashion.

As mentioned in subsection 9.1, the envelope linking effective bandwidth is about 200 Hz, leading to imperfect expansion and some speech distortion. The effects are most noticable for a deep-bass voice. However, the use of preemphasis at the transmitter significantly increases intelligibility, but lowers the overall quality due to a substantial increase in high frequency noise from the speech source. Deemphasis following expansion, minimizes this problem while maintaining a high degree of intelligibility, and virtually eliminates the most objectionable distortion components. Deemphasis also dramatically reduces the "breathing" effect of the expandor output noise during loud speech passages. Thus, the configuration which results in excellent performance when narrowband envelope linking is performed is 6 dB/octive preemphasis prior to EN, and 6 dB/octive deemphasis following expansion. Time did not permit additional experiments to be conducted to determine whether other PE/DE slopes might produce even better results. An additional bonus gained by the PE/DE operation is an additional 7.7 dB of speech SNR improvement when the channel noise is white (flat).

Having found that PE/DE improves performance with expansion, PE/DE was also tried when listening directly to EN speech. (See the discussion in subsection 9.3 with regard to the use of preemphasis in conjunction with EN speech listening.) Again, a quality increase is discerned. However, the deemphasis network also tends to significantly lower the volume level of the unvoiced segments, a result which is contrary to the reason for listening to EN speech in the first place (refer to subsection 9.3).

10.0 ENVELOPE NORMALIZATION DEMONSTRATOR

10.1 Capabilities and Features

The EN demonstration breadboard has been designed to provide all of the EN and related functions described throughout this report. Various configurations and conditions may be conveniently established by means of toggle switches and rotary controls. Figure 48 is a functional block diagram, and Figure 49 is an illustration of the front panel of the demonstrator.

The following list of functions and configuration options are provided.

(1) Choice of microphone (low-level) or high-level source inputs,
(2) AGC in/out,
(3) Preemphasis in/out,
(4) VOX in/out and VOX threshold adjustment,
(5) Choice of CCD HT or wideband 90° phase shifter,
(6) EN signal LPF in/out,
(7) Envelope LPF wide/narrow,
(8) Channel noise on/off
(9) Envelope linking on/off,
(10) EN signal expandor in/out,
(11) Speech signal listening level control,
(12) Channel noise level control.

It can be seen from Figure 48 that two speech listening outputs are provided, a standard output and an EN output. The standard output provides normal (unprocessed) speech to which noise may be added, while the EN output involves the various types of EN processing. Internal level equalization provides equivalent listening volume at both outputs irrespective to option settings. The headphone output is intended to be used with stereo headphones, and a selector switch allows listening to the standard output only, EN output only, or both outputs simultaneously in split fashion. This latter configuration permits a stereo type of subjective performance comparison.

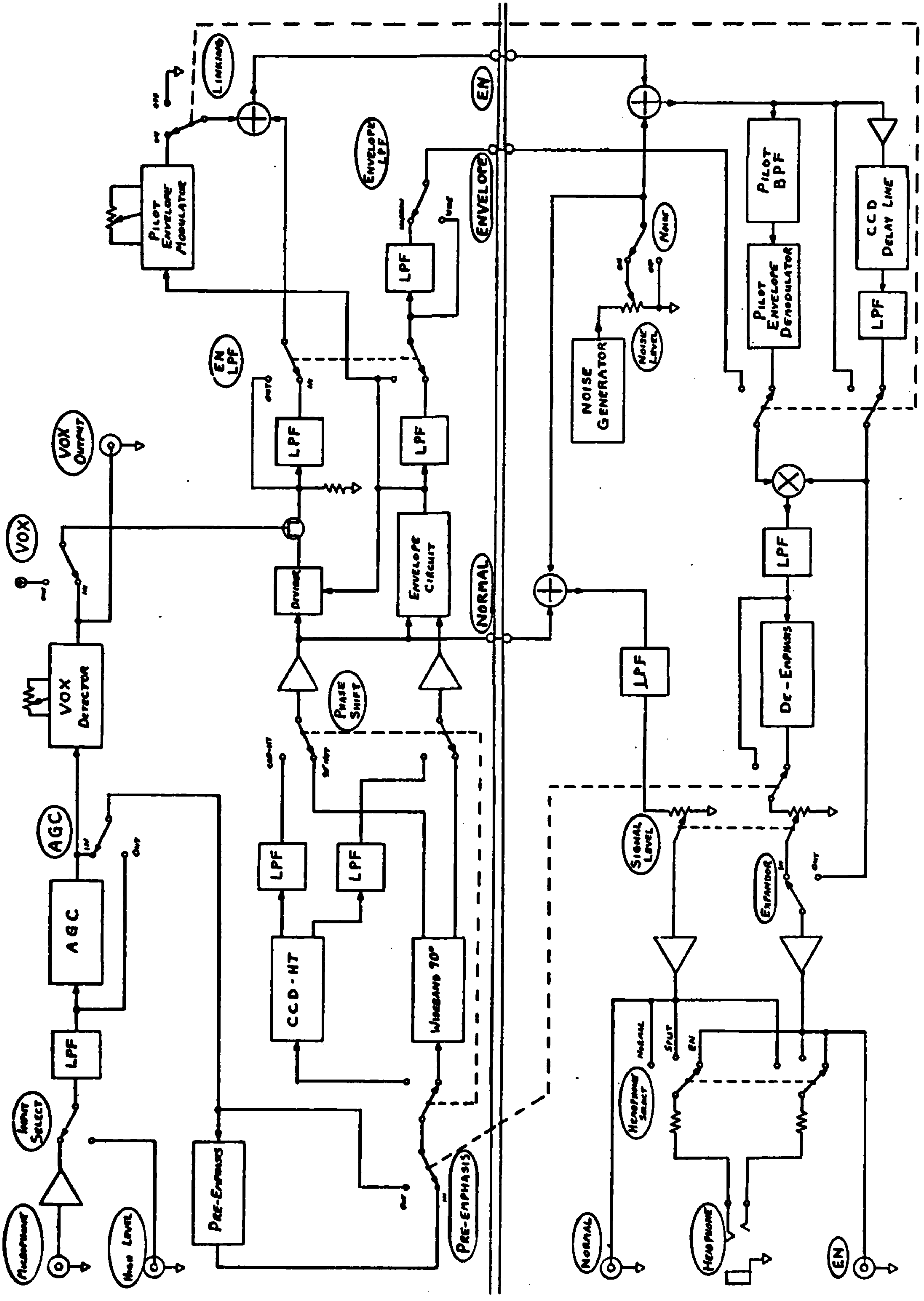


FIGURE 48 - DELIVERABLE BREADBOARD FUNTIONAL DIAGRAM

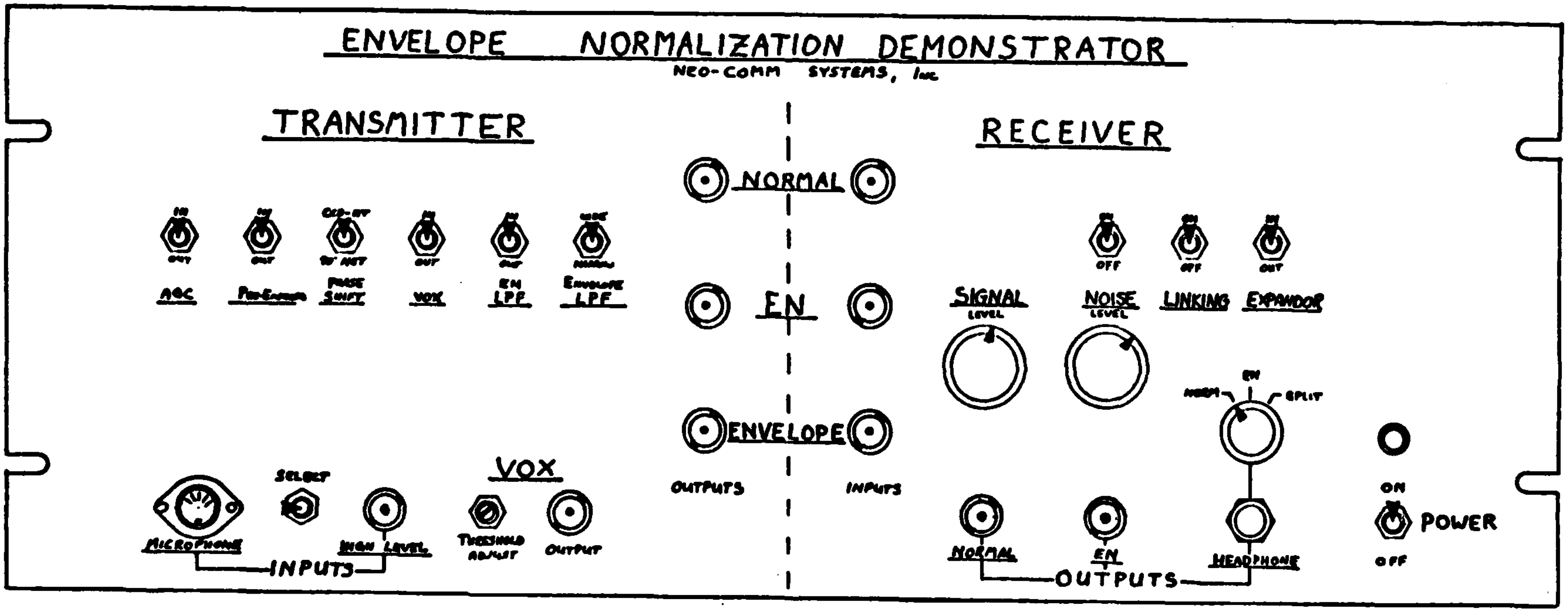


FIGURE 49 - EN DEMONSTRATION BREADBOARD FRONT PANEL

Referring to Figure 49, three transmitter outputs (labeled NORMAL, EN, and ENVELOPE) and the corresponding receiver inputs are provided. When the demonstrator is used by itself, these connectors are bridged by short cables. If the demonstrator is to be operated in conjunction with external equipment, such as a radio set, the short cables are removed, and the obvious connections made to the equipment. All three output levels may be as high as 10 V-peak depending upon options selected, so attenuation may be necessary before input to the external equipment. Corresponding amplification from external equipment to the receiver inputs may also be required.

10.2 Construction

Only the essential aspects of the EN demonstrator construction will be summarized here. The principal circuit design approaches have been outlined in Sections 5.0, 6.0, and 8.0. Detailed circuit schematics are not provided within this Report, but are supplied with the demonstrator together with a set of operating and internal adjustment procedures.

Most of the analog circuits operate from ± 15 V supplies, and the digital TTL circuits require a +5 V supply. These supplies are manufactured by Acopian Corporation. The audio output power amplifier (LM1877) uses +18 V, and the internal configuration changing relays require 12 V. A separate power supply was designed and constructed to provide these latter voltages.

The demonstrator has a standard (7" x 19") relay rack panel. All input and output connectors, with the exception of the microphone and headphone jacks, are BNC types. Mounted behind the panel is a card cage which accommodates 4.5" x 6" boards. The basic card is manufactured by Page Digital, Inc., and nine boards are used for the entire breadboard. (No attempt was made to "pack" the boards or produce a compact breadboard demonstrator.)

Portions of the assembly are covered by fiberglass panels to prevent finger access to the 120 V AC line potential. Along the top side of the card cage a row of sixteen test points permits access to various signals needed for circuit alignment. Alignment is performed by means of 10-turn trimpots located on the ends of the circuit cards. Each of the boards requiring ± 15 V is protected by IN4003 diodes to prevent component damage should the card be inadvertently inserted in the reversed direction.

11.0 CONCLUSIONS AND RECOMMENDATIONS

11.1 Reflections and Projections

In this subsection the highlights of the research program are reviewed with emphasis on those aspects for which the results were below initial hopes. Also, extensions of the basic EN technique are presented, and applications apart from voice radio systems are suggested.

Clearly, the production of essentially ideal EN speech waveforms was achieved according to the highest of expectations. Expandor performance has likewise been demonstrated to be exemplary. A big surprise was the discovery of the higher frequency content of the true speech envelope relative to the syllabic frequency range (0 Hz to 25 Hz). That speech envelope frequency components greater than 100 Hz are essential to quality EN waveform production was positively established. A fundamental requirement for inherent speech signal delay as a part of any credible EN scheme was also determined. Finally, it has been verified that the EN process does not result in significant spectrum bandwidth expansion of the original speech signal. Each of these achievements represent the sucessful attainment of major goals set for the research effort.

Probably the biggest disappointment in terms of implemented function performance is the VOX. When it was found that the zero crossing detector was not able to reliably discriminate between unvoiced speech segments and background noise, the energy detector alone was used to perform speech vs. silence discrimination. It has been established that the VOX operation becomes marginal for low speech-to-noise ratios, and that the threshold setting is critical to best performance. Therefore, additional research into VOX methods is warranted. One improvement might be made by implementing an adaptive threshold that would automatically optimize the VOX operation with changing input conditions. A basis for adaptation could be measurement of the background noise levels immediately below and above the speech band. Basing the speech vs. silence discrimination on the syllabic envelope might also prove advantageous. This technique may require formation of the envelope of the speech envelope in order to obtain reliable speech-burst beginning and end detection.

A second area where further work is justified is envelope linking. The envelope modulated pilot level, relative to the EN speech signal, is felt to be too large for an acceptable amount (3 dB or less) of expanded speech SNR degradation. Also, the sum of the EN speech and pilot signals does not result in a true constant envelope waveform, although it may appear to have a constant envelope nature when the pilot is frequency modulated because the peak value is fixed and regularly attained. Therefore, additional linking techniques studies and performance assessments are needed.

At the onset of the research program it had been speculated that a method for deriving a pseudo-speech envelope directly from the EN speech signal might be possible. No basis for such technique was found in the literature, although it had been hoped that vocoder related processing algorithms might suggest an approach. Some experiments were conducted on basic vowel sounds using octive band filtering in conjunction with envelope derivation of the resulting waveform, but no generally consistent configuration or results were discovered.

It has been reported herein that a sine-pulse approximation to an EN speech waveform appears feasible but complex. Unfortunately time and resources available to the present program were not sufficient to test this assumption. One possible advantage of the sine-pulse approximation technique is minimization of other acoustic disturbances that accompany the speech signal.

Understanding of the expandor subjective SNR improvement is more heuristic than rigorous. Hopes of devising a measurement technique apart from jury testing methods were not realized. Therefore, the subjective SNR improvement for an EN speech system cannot be precisely stated.

Finally, it had been anticipated that the EN speech processor might be tested with an actual radio system. Again, time limitations precluded such activity during the current program.

Turning briefly to uses, it is believed that applications apart from radio links can benefit from EN processing. Any system that presently employs compression and expansion is a candidate. In particular, magnetic tape recording and reproduction might be substantially improved in terms of SNR, much for the same reasons radio link SNR is increased. Recording level would always be maximum and optimum. During reproduction, tape hiss and modulation noise will be greatly reduced by the expandor during low level signal segments. Since the EN signal always has a constant envelope, preemphasis and deemphasis may be used to further reduce noise without affecting the signal dynamic range into the recording medium. For this application (especially if music is to be recorded), the envelope to the expandor should not be significantly filtered. The envelope may be recorded by placing it on a pilot well above the audio band and recording it on the same track, or recording it on a separate track using a pilot within the audio band range. It is felt that EN processing will likely prove superior in performance to existing methods, such as Dolby A, and especially Dolby B. The advantages seen for EN are that (1) it is not signal level or bandwidth dependent, (2) it represents a form of instantaneous companding, and (3) the original signal may be exactly restored at the expandor output. Another application of EN is to fiber optic communication links.

A second major use of EN is in conjunction with analog-to-digital signal conversion. Direct quantization of speech and other dynamic signals typically requires a large number of quantization levels or equivalently ADC bits. As an example, program distribution via satellite using digital audio makes use of a 15 bit ADC (and a sampling rate of 32 KHz). The majority of these bits are required to maintain linearity over the audio dynamic range. Suppose this same audio signal is EN processed to produce an EN signal and its corresponding envelope. Now, since the EN waveform has a sinusoidal nature, assume that perhaps only 6 bits are needed to adequately represent it in sampled form. Thus, the EN waveform is quantized to 6 bits at a 32 KHz rate. The envelope, on the other hand, is relatively slowly varying. Suppose that it is filtered at 2 KHz, and sampled at a 5 KHz rate using 9 bits (note: 6 bits plus 9 bits equals 15 bits). Then, the effective bit rate using EN processing becomes 6 x 32 + 9 x 5 = 237 Kbps as compared to the 480 Kbps rate produced by direct sampling. The real payoff is that two separate programs may be transmitted over a channel that previously could only support a single program. This same philosophy may be applied to any signal that has high frequency components but possesses a slowly varying envelope. Additionally, EN processing may be employed in conjunction with more sophisticated encoding techniques such as delta modulation, differential PCM, etc., in order to further reduce bit rate and/or effect economies in mechanizations.

11.2 Future Effort

Extensions of the work conducted during the present program are summarized in this subsection. Several fundamental activities that have been discussed under 11.1 require additional study. These are:

(1) Alternative and more reliable VOX algorithms,

(2) Envelope linking performance improvement,

(3) Implementation of the sine-pulse approximation,

(4) Subjective SNR improvement measurements.

In addition, measurements should be conducted using a jury of listeners to obtain articulation scores for EN speech with and without expansion, and with various types of electrical and acoustical noise present.

The performance of the EN techniques with actual radios operating both under laboratory and field conditions needs to be ascertained. Although the EN demonstration breadboard constructed for the present program is totally adequate to the goals at hand, which principally involve proof of concepts and evaluation of effects, the hardware is not particularly amendable to portable operation with field radios. Many of the circuit designs could be used within operational prototypes, while others were designed with expediency in mind and therefore should be replaced. The whole question of circuit economy vs. performance, cost, constraints, packaging/size, etc., must be addressed. In fact a set of design specifications which incorporate the radio system particulars, interfaces, and operational features should be generated.

Apart from modifications to existing radio hardware, the part that EN processing can play in future radio designs should be given serious study. It is within such radios that the full benefits of EN and related functions may be realized. In addition to the virtues of EN with linking as discussed throughout this Report, the performance of two other important receiver functions may be improved.

Receiver squelch is closely allied to the envelope information. In fact, with a good VOX and the transmission of a zero envelope level for non-speech periods, a carrier-present squelch is effectively obtained. The linking pilot can also be employed to squelch the receiver in an optimum fashion just prior to loss of carrier. When, at the transmitter, the talker terminates transmission, the system can be designed to remove the pilot (or produce some special modulation on the pilot) a short time period before turn-off of the carrier. By this means, the receiver will be able to detect and perform squelch prior to loss of carrier. In turn, the large burst of noise ("hiss-scrunch" sound) that usually follows carrier loss in typical FM receivers may be precluded from reaching the listener. A similar technique can be implemented for signal onset, together with voice delay, to prevent loss of initial speech syllables. These methods are particularly important to push-to-talk VOX'd-carrier operating modes.

A second use that may be made of the pilot is AFC, particularly if the pilot is modulated in a manner than retains a continuous subcarrier component. AFC is very important to system performance in order to minimize the effects of adjacent channel interference where tight channel spacing (small guardband) is used. For SSB type modulation, the AFC is essential to maintaining a very small frequency error (usually less than 35 Hz) that exists due to the incoherent operation of the transmitter/receiver pair.

In conclusion, the potential for EN performance increases in existing and future radio systems has been demonstrated. Also, application to other systems is suggested. Advancement of the EN technique to operational status requires further, but not inordinate, effort.

REFERENCES

1. Fagen, M. D., Editor, A History of Engineering and Science in the Bell System, The Early Years (1875-1925). Bell Telephone Laboratories, Inc., 1975.

2. Whalen, A. D., Detection of Signals in Noise, Academic Press, 1971, (pp 61-70).

3. Bedrosian, E., "The Analytic Signal Representation of Modulated Waveforms," Proceedings of the IRE, October 1982, (pp 2071-2076).

4. Dugundgi, J., "Envelopes and Pre-Envelopes of Real Waveforms," IRE Transactions on Information Theory, March 1953, (pp 53-57).

5. Pappenfus, E. W., et.al., Single Sideband Principles and Circuits, McGraw-Hill, 1964.

6. Sabin, W., "R. F. Clippers For S. S. B.," QST, July 1967.

7. Springett, J. C., and M. K. Simon, "An Analysis of the Phase Coherent-Incoherent Output of the Bandpass Limiter," IEEE Transactions on Communication Technology, February 1971.

8. Jakes, W. C., editor, Microwave Mobile Communications, John Wiley & Sons, 1974.

9. Lusignan, B., "The Use of Amplitude Compandored SSB In The Mobile Radio Bands: A Progress Report," Stanford Radioscience Laboratory, February, 1980.

10. Richards, "Transmission Performance Assessment For Telephone Network Planning," Proc. IEE, May, 1964.

11. Weaver, D. K. Jr., "Design of RC Wide-Band 90-Degree Phase-Difference Network," Proceedings of the IRE, April 1954.

12. Bedrosian, S. D., "Normalized Design of 90° Phase-Differenced Networks," IRE Transactions on Circuit Theory, June 1960.

13. Tsuchiya, T., and S. Shida, "On the Design of Broad-Band 90° Phase-Splitting Networks," IEEE Transactions on Circuits and Systems, January, 1980.

14. DeMaw, D., editor, The Radio Amateur's Handbook, 1980 Fifty-Seventh Edition, American Relay Radio League, 1979.

15. Williams, A. B., Electronic Filter Design Handbook, McGraw-Hill Book Company, 1980.

REFERENCES

16. Rabiner, L. R., and B. Gold, Theory and Application of Digital Signal Processing, Prentice-Hall, 1975.

17. Rabiner, L. R., and R. W. Schafer, "On The Behavior of Minimax FIR Digital Hilbert Transforms," BSTJ, Vol. 53, No. 2, February 1974.

18. Howes, M. J., and D. V. Morgan, Charge-Coupled Devices and Systems, John Wiley & Sons, 1979.

19. Rabiner, L. R., and R. W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, 1978.

20. Rice, S. O., "Mathematical Analysis of Random Noise," BSTJ, Vol. 24, 1945.

21. Turner, L. W., Electronic Engineer's Reference Book, Butterworth & Co., 1976.

22. Horii, Y., et. al., "A Masking Noise with Speech-Envelope Characteristics for Studying Intelligibility," The Journal of the Acoustical Society of America, Vol. 49, No. 6 (Part 2), 1971.

23. Stremler, F. G., Introduction to Communication Systems, Addison-Wesley, 1977.

24. Thomas, I. B., "The Second Formant and Speech Intelligibility," Proceedings of the National Electronics Conference," Vol. 23, 1967.

25. Mathews, M. V., et. al., "Pitch Synchronous Analysis of Voiced Sounds," Journal, Acoustical Society of America, February 1961.

26. Niederjon, R. J., and J. H. Grotelueschen, "The Enhancement of Speech Intelligibility in High Noise Levels by High-Pass Filtering Followed by Rapid Amplitude Compression," IEEE Transactions on Acoustics, Speech, and Signal Processing, August 1976.